Master's thesis

# Effects of Interaction Method's Directness on Problem Solving

Ehsan Khakifirooz

Supervisor: Professor Markku Tukiainen

ITÄ-SUOMEN YLIOPISTO

School of Computing
Computer Science

August 2015

UNIVERSITY OF EASTERN FINLAND, Faculty of Science and Forestry, Joensuu
School of Computing
Computer Science

Ehsan, Khakifirooz: Effects of interaction method`s directness on problem solving
Master's Thesis, 5519 p., 1 appendix (175 p.)
August 2015

**Abstract:** Problem solving activity is the process of delivering a problem from its initial state to the final goal state. Two main sub-activities in problem solving process are planning future state(s) of the problem and transforming the plan into action. The literature indicates that the problem solving environment can have effect on the human problem solving.

In this study, the aim is to investigate the possible effect of interaction method on a users' first person feeling (directness) and problem solving performances. The 8-puzzle game is used for studying the human problem solving. In particular, this research investigates the possibility of using the tile inter-move latency in 8-puzzle, as an indicator of the user problem solving performance. The same 8-puzzle was implemented by three different interaction methods, Touch, Gaze-augmented, and Simulated speech.

We used Gaussian mixture models as an attempt to classify the inter-move latencies, of the 8-puzzle tiles, into planning and action phases as the main sub-activities in problem solving. Manual classification was used as the ground truth of classification for two algorithms, hard k-means and a modified version of soft k-means.

As a result, the tiles inter-move latencies could be classified into the abovementioned phases and interaction method directness was found to affect the user problem solving strategy. Also, the modified version of soft k-means could provide 90% accuracy which can help to obtain the threshold between planning and action phases algorithmically.

**Computing Reviews (1998) Categories and Subject Descriptors:**

H.1.2 [User/Machine Systems]: Human factors, Human information processing. H.3.4 [Systems and Software]: Performance evaluation (efficiency and effectiveness). H.5.2 [User Interfaces]: Input devices and strategies (e.g., mouse, touchscreen), Interaction styles (e.g., commands, menus, forms, direct manipulation). I.2.8 [Problem Solving, Control Methods, and Search (F.2.2)]: Plan execution, formation, and generation

**Keywords:** Problem Solving, Planning, Directness, Manipulation Mode

# Foreword

This thesis was done at the School of Computing, University of Eastern Finland during the academic year 2014-2015, under the supervision of Professor Markku Tukiainen. The purpose of the research is to study the effect of interaction method directness on problem solving.

I want to extend my gratitude to my supervisor Professor Markku Tukiainen for providing me a chance to do the current thesis and the complete freedom during the research. I thank Tersia //Gowases for her guidance and support through all the path of the thesis research process.

I also thank people who help me on the research, especially Dr. Roman Bednarik, Dr. Ville Hautamäki, Shahram Eivazi, Hoorieh Afkari, and Mohammad Rezaei.

Finally, I am thankful to my family, friends, and specially my wonderful fiancé, Mahsa Mottaghi, for her encouragement, support and love.

This would not be possible without you all!

Joensuu, August 2015

Ehsan Khakifirooz

# List of abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| CE | Central Executive |
| DCM | Direct Concept Manipulation |
| DM | Direct Manipulation |
| DOM | Direct Object Manipulation |
| EIP | Elementary Information Process |
| GOMS | Goals, Operators, Methods and Selection rules |
| HCI | Human-Computer Interaction |
| IPM | Information Processing Model |
| IPS | Information Processing System |
| LTM | Long-Term Memory |
| PL | Phonological Loop |
| RDCM | Reflective Direct Concept Manipulation |
| SDA | Sense-Decide-Act cycle |
| SM | Sensory Memory |
| STM | Short-Term Memory |
| TOTE | Test-Operate-Test-Exit |
| UEF | University of Eastern Finland |
| VSP | Visuo-Spatial Sketchpad |
| WM | Working Memory |

# Contents

# List of Figures



### Appendix

# List of Tables



**Appendix**

# 1 Introduction

> "A problem arises when a living creature has a goal but does not know how this goal is to be reached. Whenever one cannot go from the given situation to the desired situation simply by action, then there has to be a resource of thinking… Such thinking has the task of devising some action which may mediate between the existing and the desired situation" (Duncker, 1945, p. 1).

During nineteenth century, the initial studies in the problem solving area contained mostly the psychological research on the nature of thinking and problem solving. The Gestalt psychologists believed that problems solving should cause insight into the problem, in which subjects is intended to explore the problem and find the solution element, in contrast to behaviorists who believed that solving a problem is based on trial and error (Dunbar, 1998).

The focus on the problem solving research changed to a systematic view during the second half of the 1950s by the study of Newell and Simon on the human problem solving (Dunbar, 1998; Ohlsson, 2012). Analysis of Human behavior including goals and plans has resulted in valuable information in the problem solving domain (Newell & Simon, 1972). Subgoals or actions are the elements of plan structure inside human memory (Robertson & Black, 1986).

Newell and Simon (1972) explained problem solving in relation with both the problem solver and environment. The effect of environment on problem solving has been the interest of many researches, such as the effect of external environment context on human problem solving strategy (O'Hara & Payne, 1998), and modeling human interaction with environment (Norman, 1986). One of the commonly used aspects in problem solving research area is manipulation of the external environment in terms of human feeling of directness and measuring the effects on different human mind processes. Hutchins et al. (1985) described the feeling of directness by the distance, which is the amount of user opportunity to access context in computer instantly, and engagement, which is the feeling of user as the real actor of interaction.

Norman (2002) defined the modes of manipulation as the methods of doing a task with computers. Many human computer interaction devices are using command-driven interface in which user's plans are being issued to the computer by submitting command(s), for instance entering command by typing a word on a keyboard. The advancement of user interface technologies has caused one of the major changes, direct manipulation. The aim of this manipulation is to make the user intention closer with the task than before. Direct manipulation enables humans to have opportunistic and incremental planning (Hayes & Hayes, 1979) during the problem solving task with low loads of mental efforts and planning. Whereas, the command- driven style is oriented to result in action(s) with higher planning and mental efforts. In general, these two manipulation modes are being researched from different aspects, such as controlled search and automatic detection (Shiffrin & Schneider, 1977), complete and incremental planning (Hayes & Hayes, 1979), plans and situated actions (Suchman, 1987), and plan-based and display-based strategies (O`Hara & Payne, 1998).

O`Hara and Payne (1998) showed that direct-manipulation can effect different properties of problem on subject shift between planned and situated action. Such effect of manipulation on subject problem solving behavior includes research with different user interface technologies, for example touch screen systems (Kieras, 2001), gaze-augmented systems (Bednarik et al., 2009), and tangible user interface (Manches et al., 2009).

The effects of manipulation mode on the user problem solving created the motivation for this research. This research investigates the utilization of different levels of directness on human computer interactions for a common problem solving task, and studies how they influence planning and action of the problem solver. The 8-puzzle is utilized as the problem solving task, because of its clear reflection of subject`s behavior.

The study involves an analysis of different measurements of subject behavior, and the effect of manipulation of human computer interaction on subject problem solving behavior. The measurements are referenced from previous studies conducted by Ericsson (1974b) on the 8-puzzle planning process study, however with a different focus.

### 1.1 Research questions

The aim the research is to answer the following questions on user problem solving during interaction with different manipulations of human computer interaction:

1) Can inter-move latencies be used to measure problem solving performance for the 8-tile puzzle game?
2) Does the amount of interaction method directness affect inter-move latency?
3) What are the effects of problem solving processes on inter-move latency? (i.e. to use latency data to identify different phases of problem solving)

### 1.2 Research method

The used research method for this research is quantitative research method. In the first phase of analysis, the achieved raw quantitative data from experiment was processed statistically and analyzed using the ANOVA method. Then the Gaussian mixture models analysis was used to explore the phases of problem solving. The obtained results from both phases of analysis are used to test the research questions.

### 1.3 Thesis structure

The thesis consists of five chapters as follows:

1) Introduction: Provides a glimpse of the background of the current studies, as well as, a brief discussion on the performed study.
2) Cognition and problem solving: Includes a review of the previous researches in the area of problem solving and the related studies in the Human Computer Interaction (HCI) domain.
3) Experiment: Reviewing the design and settings of experiment, and the process of the data collection.
4) Results and Discussion: Providing the method of analysis and the discussion of results.
5) Conclusion: Concluding the study and providing a further study point.

# 2 Cognition and problem solving

Cognition is defined as all mental functions which convert, reduce, expand, store, retrieve, and utilize sensory data (Neisser, 2014, p. 9). Cognitive psychology is a specialty of psychology which focuses on realizing and knowing, and describes the mental functions in details (Groome, 2013, p. 3).

Problem solving is a complex concept that humans encounter in their everyday life. In general, it can be defined as the analysis and transformation of information to reach a determined goal (Lovett, 2002). By the progress of cognitive psychology, the research in learning and problem solving has increased, and it involves identification of mental stages along problem solving processes (Hardin, 2003).

Human-Computer Interaction (HCI) is the study of the interaction between humans and computer technology. Cognitive psychology is one of the main contributors within the HCI research domain, which provides psychological basis creating models which makes a view of human performance (Giacoppo, 2001).

In the following sub-sections the mental processes, including the problem solving process as a mental process, and the HCI related topics in problem solving have been discussed.

## 2.1 Mental processes

Mental processes are the cognitive functions which act on the representation(s) we make from our knowledge of the world around us (Winn & Snyder, 1996). It can be exemplified by attention, memory, reasoning, problem solving, and decision making (Goldstein, 2008, p. 2; Luine, 2014). Cognitive psychology is a study of the mental processes and it allows us to describe and categorize them.

In general terms, cognitive psychology includes two assumptions: the first is that cognition of humans can be explained by scientific methods which help to explore individual parts of mental processes, and the second is that these internal mental processes can be explained by applying methods of information processing models (Lu & Dosher, 2007).

Information processing integrates the human problem solving mechanism into a computational model (Laurillard, 1997). Moreover, the Information Processing Model (IPM), models the human brain mechanism by including attention, as the input data function, working memory, for processing data, and long-term memory, for storing data for future utilization (Meyer, 2004).

Using the abovementioned assumptions, some of the mental processes are briefly explained in the next sub-sections.

### 2.1.1 Attention

Attention is the process of assigning restricted mental resources (e.g. auditory, cognitive, visual, and motor) to cognitive processes (Anderson, 2000, p. 104; Sternberg & Sternberg, 2012, p. 137). It includes four main functions, which are defined in short as follows:

1) **Signal detection and vigilance:** Monitoring the observation (Vigilance) to detect a particular target stimulus of interest (Signal)
2) **Search:** Detecting a particular target stimulus of interest (Signal)
3) **Selective attention:** Neglecting some stimuli, and highlighting target stimulus of interest
4) **Divided attention:** Sharing attention resources to between some tasks

### 2.1.2 Memory

Memory is the process to store and restore sensed/sensory information. It has been systematically studied since the first discussions on the forms of the memory in between 1960s and 1970s (Schacter & Tulving, 1994; Tulving, 1995) and can be classified into Sensory Memory (SM), Working Memory (WM), Short-Term Memory (STM), and Long-Term Memory (LTM) (Goldstein, 2008, p. 143; Sternberg & Sternberg, 2012, p. 194). Atkinson and Shiffrin (1968) provided the first model of memory “modal model” which considers memory as a set of stores (also known as memories) working together. Figure 2.1 illustrates the model.

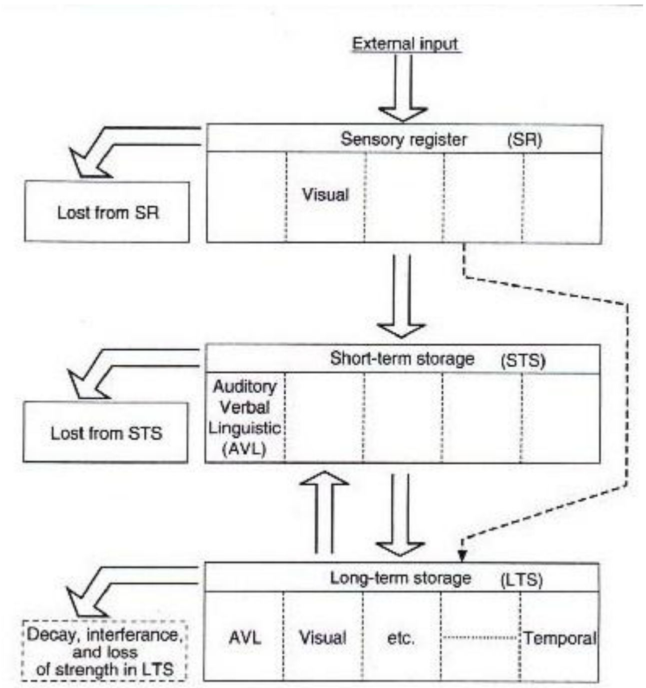



**Figure 2.1. The modal memory model (Atkinson & Shiffrin, 1968)**

SM is the entrance point for information inside human memory, which contains the exact copy of the sensed data (Coon & Mitterer, 2010, p. 253). Sperling (1960) discovered SM by studying iconic store, which resulted in the subject information recall from 4 to 5 symbols by the decay time of less than one second. By further researches, Darvin et al. (1972) repeated the Sperling research method on auditory stimuli and found echotic store as the auditory stimuli data store, and Shih et al. (2009) inroduced haptic store as for haptically acquired information.

STM receives a big portion of the information through selective attention on sensory memory (Cowan, 1988). The STM can store information in a short period and with a limited amount of space, which is 7±2 chunks of information due to the limited number of memory slots (Miller, 1956).

WM is a mind system which brings the ability to maintain goals, ongoing processes and future actions (Henry, 2011, p. 24). The difference between WM and STM is quite narrow. One of the differences is that WM includes the processing unit to manipulate the information inside the short memory that is placed in WM (Cowan, 2008). Miller et al. (1960) have considered it as the unit for planning and behavior, which can even have more than one plan in the inner memory (p.65).

Baddeley (1986) introduced the organization of WM, which consists of three components, Visuospatial Sketchpad (VSP), Central Executive (CE), and Phonological Loop (PL). CE has the role of performing cognitive processes and controlling and information transaction on various passive systems (also known as "slave systems"). For example, two passive systems are VSP, which is used for visual and spatial information, and PL, which is used for acoustic or speech content. The PL includes two components: a phonological store (inner ear) which keeps speech-based information for 1-2 seconds, and an articulacy loop (inner voice) which helps to rehearse and maintain an amount of verbal information from the phonological store in a fixed duration (Baddeley, 2000). CE also carries out cognitive tasks like arithmetic operations and problem solving (McLeod, 2012).

Baddeley (2000) revised the model with a new component "episodic buffer" which represents a buffer memory for communication between central executive component and long-term memory, an extra amount of memory for central executive operations, and an integrated memory for all components. Figure 2.2 includes the revised model.

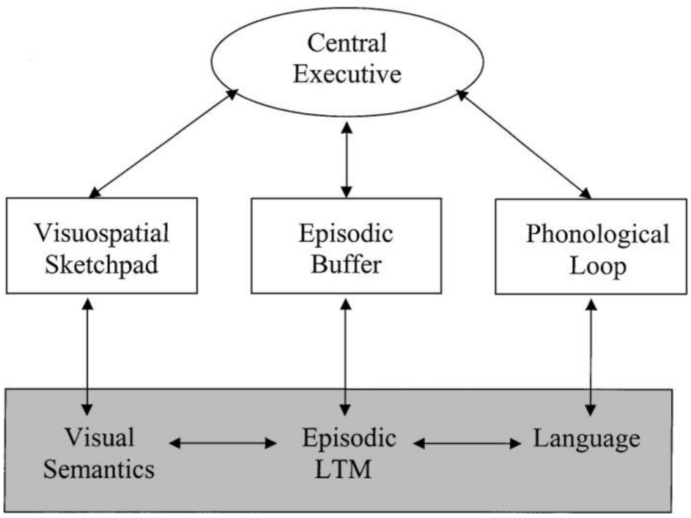



**Figure 2.2. General model of working memory (Baddeley, 2000); gray section includes the Long-term Knowledge systems which provide the communication between components of working memory**

McLeod (2012) refers to CE as a unit which is processing (i.e. combining) information from sensory components (the phonological loop and the visuospatial sketchpad), and is writing on information being kept inside a large database (LTM).

The LTM can keep information by longer time, or perhaps with unlimited time (Richardson-Klavehn & Bjork, 2002).

### 2.1.3 Problem solving

Problem solving is the process of conquering problem obstacles which block the path to a solution (Sternberg & Sternberg, 2012, p. 443). Problem solving is presented by searching the space of problem states, which provide the problem condition regarding to solution, initial state, which presents the first situation of problem, and the goal state, that can be reached by taking some intermediate states from the initial state (Anderson, 2000, p. 242).

The first theories on problem solving (Thorndike, 1898; Kohler, 1927; Tolman, 1932; Guthrie, 1952) were mainly all about the learning process during problem solving, which did not address a systematic analysis of the problem solving process itself (Anderson, 1993). Miller et al. (1960) provided the analysis of problem solving by utilizing information processing theory, which is known as TOTE (Test-Operate-Test-Exit). TOTE can be considered as the elementary unit of human behavior which includes two phases:

1) Test phase: It includes the information for checking the incongruity or congruity of current state (received information by organism) with the expected state.
2) Operation phase: It is an effort to produce a plan to produce the answer that test phase is looking for.

Figure 2.3 shows the structure of a TOTE unit. The TOTE units were referred to as the constructor of human general behavior during interaction with the environment (House & House, 1987). In general, TOTE became the foundation for the many other theories in problem solving domain (Adams, 2009, p.249).

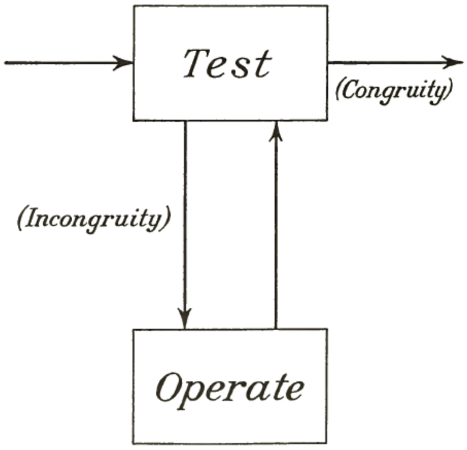



**Figure 2.3. The TOTE unit (Miller et al, 1960); it can contain more tests than action by a hierarchy of TOTE units inside the operation phase according to the complexity of planning**

Newell and Simon (1972) describe human problem solving as a cognitive model for problem solving including two basic components: the environment including the task, and the Information Processing System (IPS). Humans are considered as IPS systems which mainly includes four components: receptors and effectors which are in interaction with the environment, a memory, and a processor which processes data with the help of the other three components. A general view of the IPS is shown in Figure 2.4.

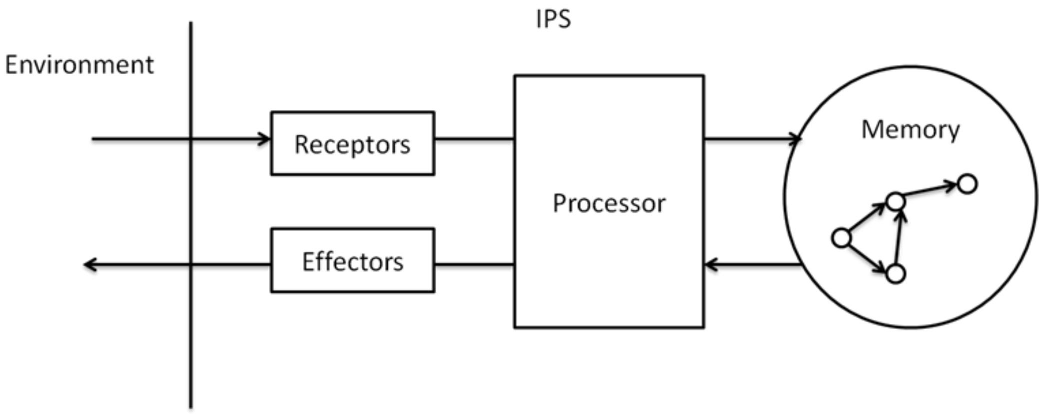



**Figure 2.4. General structure of an IPS (Newell and Simon, 1972)**

As the model contains an IPS, the data unit which is used inside the processing activity is called a symbol. Symbol structure types represent an object, that carries semantic data, or programs, that provides the operations, or methods which can supply information from the symbol structures or can manipulate them (Newell & Simon, 1972, p. 23; Smith, 1994, p. 68). By considering Elementary Information Process (EIP) as a process having certain input and output symbol structures, each component in the system can be defined as follows (Newell & Simon, 1972, p. 20):

1) The Memory: The unit which is utilized to keep and fetch tokens of symbols, which are related to each other.
2) Processor: The unit which keeps a series of EIPs that are being defined as logical rules (production rules), stores tokens of symbols of each EIP in an embedded STM, and determines the execution order of EIP`s by its interpreter unit.
3) Effectors and Receptors: The units which provide the interface for the interaction of IPS with the task environment.

IPS's behavior can be a set of rules and conditions which explain the order of EIPs for execution based the context. The rules can be called a program, and can be implemented in a programming language to describe human problem solving. The task environment is the environment which includes the goal, problem, or task which the subject is intended to be in contact with it. The behavior that subject provides in problem solving conditions, which is known as adaptive behavior, is towards the goal by considering the problem solving environment and its conditions (Newell & Simon, 1972, p. 53). Hutchins (2000) has mentioned the environment as the external (material or environmental) structure of the problem, which can be a computational medium.

Fundamentally, we should consider the external environment, as the container of the external representation, separated from the internal representation of problem, as the symbol structure which can provide the gathered data when interaction of the stimulus with the environment has been transformed in a different format inside the memory. Internal representation is the mind`s mental image of the problem, whereas the external representation is real world, physical symbols of the problem (Zhang, 1991). Newel and Simon (1972) have mentioned this internal representation as the problem space which has its structure affected by the structure of the environment (p. 59).

The subject initially provides the goals, rules, constraints of the problem, and other problem components into an internal representation which includes initial state, intermediate states, goal state, and rules. Based on having the definition of internal representation, Newel (1979) defines planning as the abstraction of the current state of the problem in environment and continue solving the abstracted problem by applying the found solution path on the unabstracted problem. In general, decision cycle (planning) can be led into developing some sub-goals and constructing state space for each sub-goal and choosing the appropriate sub-goal to reach. The sub-goals can cause the generation of new representation for the related solution (Smith, 1994, p. 69).

Newel (1990) altered his perspective about the IPS model and proposed a unified theory of cognition by an architecture (model) called Soar. The developments in the newer version of the initial IPS model were dividing memory into LTM and WM, encoding both declarative knowledge and programs as production rules, and learning by producing new production rules that relate the taken path from a certain state to a goal (or sub-goal) state, which is known as chunking process. Figure 2.5 shows the Soar architecture.

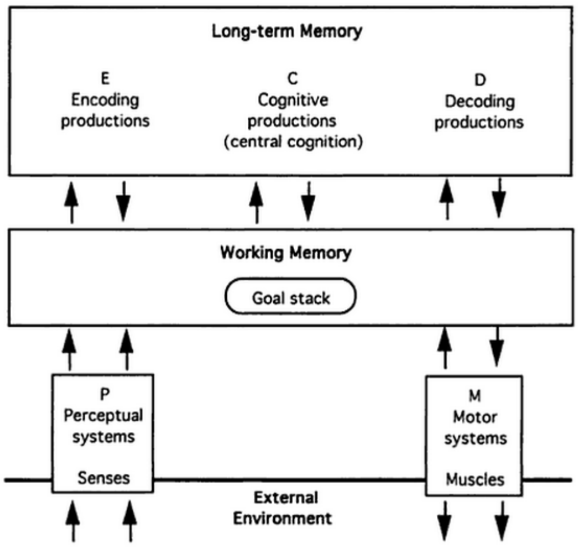



**Figure 2.5. Soar cognitive architecture (Newell, 1990)**

Soar is known as the main model of human cognition. Moreover, Newel (1990) refers P-E-A-C-T-I-D-M scheme as a process which contains the main functions, necessary to do a task. In general, P-E-A-C-T-I-D-M is a presentation of basic processing stages of Soar control loop, which is from perception (P) to encoding (E) to attending (A) to comprehension (C) to tasking (T) to intending (I) to decoding (D) to motor action (M).

Wray and Jones (2005) considered Soar cognitive model as Agent architecture in the Artificial Intelligence (AI) domain. Soar's decision cycle is based on fundamental perception-action cycle, three steps of Sense-Decide-Act cycle (SDA), which is illustrated in Figure 2.6. From the left side the abstract SDA can be viewed as Soar decision cycle, middle diagram, and the detailed SDA of Soar on the right side.

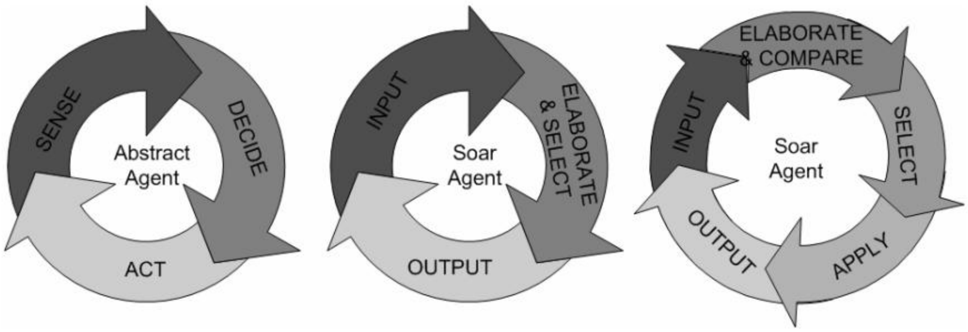



**Figure 2.6. Soar's decision cycle (Wray & Jones, 2005); from left to right: SDA in abstract agent, Soar agent SDA, and a detailed view of the Soar agent SDA**

A Soar agent receives input information during the Input phase, and its execution of commands on the problem inside the environment is happening during the output phase. Between the considered phases, the decision cycle is happening in three distinct phase: elaboration and compare phase, which includes the agent production rules (operators) execution and comparing the results with the goal, select phase, which includes the selection of operator based on preferences of system, apply phase, which is loading the operator`s production rule into the working memory and applying it on the knowledge. The plans are being created by having the operators as the fundamental elements which execute actions based on the plan steps (Wray & Jones, 2005).

## 2.2 HCI and problem solving

Congnitive architechture is a method of integrating the knowledge about human cognition and performance (Byrne, 2005). In an attempt Card et al. (1986) developed a cognitive architechture, the Model Human Processor, which describes the way a human behaves during the process of interaction using a determined computer system and set of tasks. This type of description provided a systematic view and a perspective of information processing system for prediction of gross system behaviour.

The Model Human Processor consists of the perceptual system, the motor system, and the cognitive system, by having each system with dedicated memories and processors. The considered model can be explained by:

1) The organization of memories and processors.
2) Principles inside the system (for a detailed explaination please refer to Card et al., 1986).

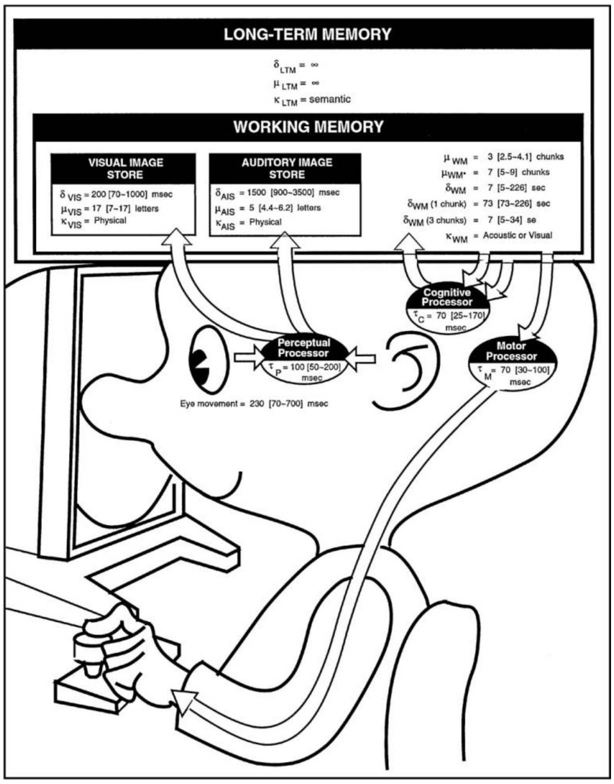



**Figure 2.7. The organization of memories and processors (Card et al., 1986); the π is the storage capacity, the δ is decay constant, and the κ is the main code type**

Figure 2.7 shows the organization of components (memories and processors) and the communication among them. The perceptual processor interprets data about the physical world detected by the body`s sensory systems into an internal representation being stored inside the perceptual section of the working memory. Shortly after gathering the sensory information, a part of the data would be brought into the working memory data beside the related retrieved data from long-term memory. The information inside Working Memory is being processed by Cognitive processor, later, it will be translated into action by activating the voluntary muscles of body. Therefore, the reaction-time, time from perception to action, is:

$$T_P + T_C + T_M$$

Where $T_P$, $T_C$, and $T_M$ are perception time, decision making time, and motor time, respectively. According to Variable Perceptual Processor Rate Principal, the $T_P$ value would be different according to conditions of the physical world (Card et al., 1986). Moreover, according to Variable Cognitive Processor Rate Principal, the $T_C$ value would be shortened by exercise, task pacing, higher attempt, or less accuracy.

The $T_C$ principally consists of a series of recognizes-act cycles, that each cycle is a set of parallel actions related to long-term memory information manipulation or retrieval (recognize) and serial modification of the working memory data (act). Additionally, plans, providing a sequence of tasks, and other kinds of possible behaviours are being provided by a set of recognize-act cycles.

According to the rationality principle, different user behavior can be indicated by means of reaction-time, which denotes that human behavior is not only the result of the human internal system, however, it is also caused by the complexity of task environment which includes goals and goal search activity. Therefore, the environment and the conditions of problem solving activity play a key role in user behavior.

Card et al. (1986) also defined GOMS (Goals, Operators, Methods and Selection rules) concept for modeling user interaction with a computer on a problem solving level. Newell (1990) Soar’s cognitive architecture is quite close to IPS model and the GOMS architecture, as Newell was an author/co-author of both previous researches.

The GOMS model uses production rules, which uses actions and condition, to predict the users' actions under a range of situations (e.g. learning, working memory processing, and taking action). Since goals can contribute in both actions and conditions, goals and subgoals can be determined by using a collection of basic operations. As a result, the computed time values beside the collection of basic operations can aid in the prediction of task time consumption (Smith, 1994, p. 86).

The models introduced by Card et al. (1986) are computational design tools for HCI (Norman, 2002, p. 221) which provide approximate and quantitative theoretical action model for task analysis to be applied on the real problem of user interface (Newell and Card, 1985).

Norman (1984) considered the process of interaction of a human with a computer system, as the subject's attempts to satisfy intention, by a non-computational model in four different stages of activities, as follows:

1) Intention: forming the understanding about the current state of system and the possibilities for the next state.
2) Selection: Selecting an individual action or a sequence of actions based on the formed intention.
3) Execution: Executing the action(s) on the computer.
4) Evaluation: Feedback about the new state of the system.

In general, execution stage can be done in two ways: 1) the usual method of running commands in computer system, and 2) pointing on the display to select the command to run. The supporters of the former claim it is easy by execution aspects, and the supporters of the later claim it is easy by selection aspects (Norman, 1984).

In a later study, Norman (1986) provided an action theory which clarified the boundary between user and environment inside the stages of activities. The theory included seven stages of activities as follows:

1) Determining the goals.
2) Forming the intention.
3) Determining the sequence of actions.
4) Executing the sequence of action(s).
5) System state perceiving.

6) System state interpretation.
7) System state evaluation regarding the Goals and Intentions.

Norman's revised action theory considered the difference between terms related to human goals, also known as psychological terms, and the terms related to system interaction method and states, also known as physical terms, by creating Gulf of Execution and the Gulf of Evaluation. Each gulf can be crossed by a bridge which will be discussed in details. Figure 2.8 shows the gulfs and bridges of execution and evaluation.

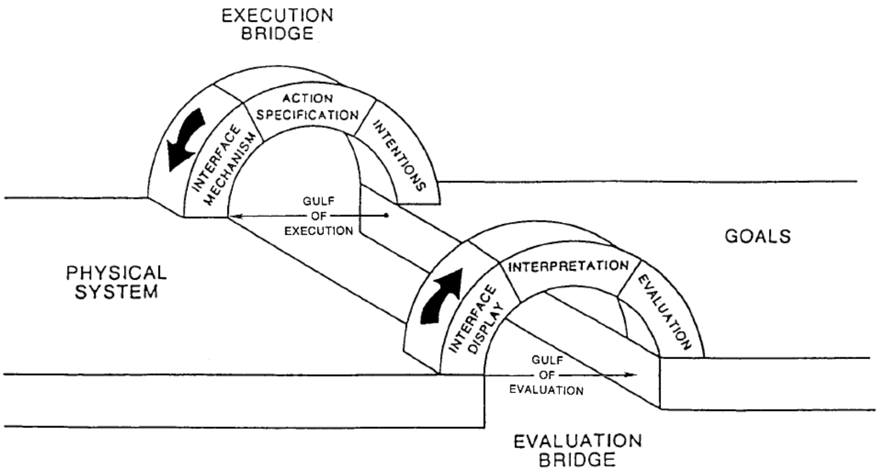



**Figure 2.8. The gulfs and bridges of execution and evaluation (Norman, 1986)**

The space from Goals to Physical System (The Gulf of Execution) is bridged by having a sequence of 4 steps:

1) Forming the Intention: The step which converts the internal representation of a system (the way user thinks about the system) into its external representation (the way the system is) (Moran, 1983)
2) Planning the sequence of actions: The step of obtaining a sequence of actions from goals of formed intention.
3) Contacting with the user interface: The step of executing the planned sequence of actions.
4) Interaction with the physical system.

After the first step is taken, the second step includes mappings between intentions and physical actions, physical actions and the physical state of problem, physical state of problem and the user goals and intentions. It generally means planning for the next state of problem in the terms of goals and intention. The third step is a set of actions to be executed in order, which can be according to the type of the user interface being used for the interaction. Therefore, the level of obtaining user goals can be affected by the user interface effect on the choice of actions.

The space from Physical to System Goals (The Gulf of Evaluation) is bridged by having a sequence of 4 steps:

1) Displaying the output of current state.
2) Interface display: The step which provides the result of the executed actions on the interface and the interface sensory data is received by user to provide the user perception.
3) Interpretation: Process of the received perceptual data.
4) Evaluation: Comparing the interpretation result with the previous user`s intentions and expected goals.

The gulf of execution can be considered as the difficulty of acting in the external environment, and the gulf of evaluation can be considered as the difficulty of evaluating the current state of the external environment (Norman, 1991).

## 2.3 Directness of interaction

Directness of interaction can be defined by Direct Manipulation (DM) concept, that was proposed by Shneiderman (1982, 1983) as a form of HCI where user interfaces include the visualization of content for the purpose of continuous object representation and manipulation against complicated syntax, and quick reversible operations with quick system output.

What makes an interaction with direct manipulation property, is discussed by Hutchins et al. (1985) with defining two phenomena as follows:

1) Distance: The space between the user`s intention and the physical system requirements.
2) Engagement (aka Direct engagement): The feeling for the user to have the first person role in the manipulation.

The distance can be minimized by reducing the mental effort a subject is going to have through interacting with a system. Mental effort is the directly related to the evaluation and execution gulfs, which can be bridged effectively with a better system interface design (Hutchins et al., 1985). The direct engagement can be achieved by the amount of user`s perceived locus of control of action inside the system (Laurel, 1986). Frohlich (1997) exemplifies users using speech interaction with an indirect engagement and second-person feeling, and users using touch-screen interaction with a direct engagement and first-person feeling. Therefore, the complexity level of computer input system can affect how well a user chooses the actions and matching between user intention and the system state (Norman, 1986).

DM can be also considered as an interaction style. Norman (2002) classified interaction styles into two modes of manipulation as follows:

1) Direct manipulation mode: User does the task directly
2) Command mode: User orders computer to do the task

The Command mode has been mentioned as a third person mode, which involves less engagement of user with the interaction. The direct mode, on the other hand is thought to be a first person mode, where the user is completely engaged with the interaction (Norman, 2002, p. 184).

## 2.4 Directness and problem solving

Hayes and Broadbent (1988) defined two different modes of learning, Selective mode (S-mode) as learning by conservatively processing the perceptual context inside the working memory, and Unselective mode (U-mode) as learning with the aid of external environment context. Svendson (1991) researched the effects of interaction manipulation on mode of learning by implementing the Tower of Hanoi puzzle in previously discussed computer utilization modes. Svendson (1991) reported direct manipulation mode was oriented to U-mode learning, whereas command mode was utilizing S-mode learning by having higher trial time, less number of trials, and fewer errors.

Ericsson (1974) provided a study on 8-puzzle and problem solving. The purpose of using the 8-puzzle was the broad space of state it includes and including significant load of problem-solving process. The 8-puzzles with short minimal path solution indicated less cognitive activity by less time to solution in compare with the other puzzles (Ericsson, 1974a), and the subjects showed less of cognitive effort by including more action selection activity and less planning after gaining experience in solving 8-puzzles (Ericsson, 1974b). As a result, strategies with more planning required higher mental effort and had more inter-move latencies.

O'Hara and Payne (1994, 1998) extended the study of Ericsson on 8-puzzle by studying the effect of different properties of problem on subject`s shift between planned and situated action. For this purpose, the operator implementation cost (cost of doing a single move) was manipulated. Subjects who were using the low cost condition interface (direct manipulation mode) showed a less planful strategy which was containing searching the solution path on display (display-based planning strategy) by doing trials in a short time, low inter-move latency, and having more error actions. In contrast, subjects who were using the high cost condition interface (command mode) showed a more planful strategy, that search paths were processed and evaluated mentally.

Trial and error behavior during problem solving has been reported as using the external representation for off-loading cognitive work onto the environment by using epistemic actions (Kirsh & Maglio, 1994) to change the environment in order to decrease the remained cognitive work. Kirsh and Maglio observed user during Tetris game preferred to manipulate the physical parts instead of mentally providing a solution and execute it.

In addition, O'Hara and Payne (1994, 1998) explained the higher planning in the high cost condition as the noncorresponding mapping between subject`s internal representation and the states represented inside the external display. On the contrary, subjects who were using in the low cost condition should have a much closer tracking between the internal and external representations.

O'Hara and Payne (1999) explored the effect of user interaction lockout, which is the interaction response time to the user action, on the planning and action by increasing the duration of time to perform an undo move in the slide-jump puzzle or putting a delay for providing the next move inside the 8-puzzle. The results were the same as their previous studies, and in the implementation with low cost, display-based planning was seen.

Due to the development of the technologies during the recent decades, research in the field of testing modality effect on the problem solving have managed to utilize the newer user interfaces. Kieras et al. (2001) applied the EPIC architecture, which a version of the Model Human Processor (Card et al., 1986), to compare the user performance in a computer visual game by two different manipulations, keypad (indirect engagement) and touchscreen (direct engagement). The subjects with touch screen interface provided a narrower gulf of execution which shown a better performance in compare with the keypad interface in the terms of easy processing of response selection and having the chance to parallelize the perceptual and motor processing with each other.

Sedig et al. (2001) have studied the effect of the interface directness on reflective cognition and concept learning by utilizing the tangrams puzzle in three different implementations with computer mouse, Direct Object Manipulation (DOM), Direct Concept Manipulation (DCM), and Reflective Direct Concept Manipulation (RDCM). Subject who used the more direct manipulation (DOM) indicated more solved puzzles and majority of them believed that they have had less thinking (planning) and more guessing (immediate action) in compare with the groups who used more detailed and command mode oriented implementations.

Manches et al. (2009) used the notion of manipulation modes in the tangible technologies design by studying the effect of physical artifacts in children's numerical problem solving. Tangible user interface design is the combination of physical representation, manipulating digitalized data, and making interaction between physical artifacts and computational system (Hornecker, 2006). Manches (2009) provided the numerical problem solving in two forms as physical condition and computer simulation, as virtual condition. As a result, the wide range of states and actions in the physical condition resulted in more learning, and children discovered better strategies by using trial-and-error actions.

Bednarik et al. (2009) studied the effect of interaction modality on user problem solving strategies, performance, and experience. The experiment participants solved

8-puzzle game in one of the three implemented interactions in the study, dwell-time selection method, gaze-augmented selection, or computer mouse selection method. The interactions were analyzed with different measurements, such as completion time, number of moves and moves per minute. The amount of mental effort for submitting a command in gaze-augmented interaction was less than the other interactions, which allowed users to spend more on planning and make better plans. In the case of mouse interaction, users performed their planning by doing trial-and-error activity on the external representation. Users who were using the dwell-time method faced with more correction of their intention; however they did not show any performance difference with users who were using computer mouse interaction. In summary, users in gaze-augmented group felt more engaged with the interaction and provided better results in completion time, number of moves to solution and number of moves per minute, in addition, provided less error during problem solving.

## 2.5 Summary

The complexity of internal representation is not only the result of human behavior, however, it is also caused by complexity of problem solving environment (Zhang & Norman, 1994). In different external environments subjects can use different problem solving strategies which can result in different use of cognitive processes. Working memory with a load of planning, or the off-loading the cognitive activity to environment of problem solving process can be done. Subjects can have a long sequence of actions as a plan or can be more oriented to display-based actions as their planning activity. Hutchins et al. (1985) considers the rapid feedback of on-display planning as a support of direct acting feeling on external representation and providing the perceptual resources as evaluator of actions.

Consequently, users provide a behavior which is adapted to the problem solving environment. Many theorists view this behavior as adaptive problem solving behavior, which is evaluated by studying if the system can provide the efficient use of the available information, if it can help the user to reach the goal state, and if it allows the user to fulfill the requirements with a changing environment (Sternberg & Ruzgis, 1994, p. 107).

In order to study the adapted user problem solving behavior,  the amount of interaction method`s directness can be modified and manipulation modes can be regarded as the key principle. The literature review included some relevant research on the modifications of directness and their respective results on the user strategy for solving a problem. In the following chapter we investigate this concept in details through using user interaction with variations of directness, and their affects of user problem solving behavior and strategy.

# 3 Experiment

## 3.1 Introduction

In this experiment, the 8-puzzle has been used to test the three stated hypothesis of the research through quantitative measurements of the human problem solving process.

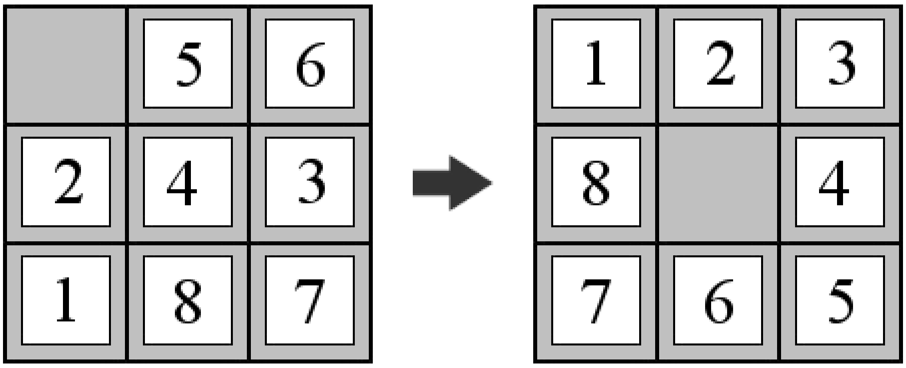



**Figure 3.1. The 8-puzzle; an example initial configuration with its goal configuration**

The 8-puzzle (see Figure 3.1) is a tile rearrangement puzzle game that consists of eight square tiles (including one empty cell) arranged in a three by three frame. Different permutations of tiles provide the different states of the puzzle, and there are so many possible initial configurations for problem solving (Ericsson, 1974a).

The aim of the 8-puzzle is to reach a goal configuration from the given initial configuration by sliding the tiles into the orthogonally adjacent empty cell (Reinefeld, 1993). The motivation for studying this type of puzzle in the study was the inclusion of broad search space which brings enough difficulty for subjects (O’Hara, 1998).

The experiment is performed to test the changes of user planning and performance through interacting with different user interaction methods. Following the modes of manipulation defined by Norman (1986), each user interaction method provides a different degree of directness for the interaction.

Hutchins et al. (1985) considers DM mode as a way to feel more involved with the world of objects, instead of having access by intermediaries. The DM mode, on a micro level, can correspond to have first person feeling by using modern tool-based interfaces (e.g. mouse), which are distinguished from old interfaces (e.g. keyboard) that bring command-mode user interface with third person feeling (Heeter, 1991).

In this study, the degree of directness for each interaction method is modified by using different interfaces. The following section involves the detailed description of the experiment.

## 3.2 Method

### 3.2.1 Subjects

A total of 24 students (5 female, 19 male, mean age = 28, SD = 4.17) from the University of Eastern Finland, Joensuu campus, volunteered to take part in the experiment and were awarded one course credit for participation. Sixteen of the subjects reported having prior experience with sliding tile puzzles, this includes the 8-puzzle, a subset of  commonly known 15-puzzle. Data from 6 subjects were excluded as they either, failed to complete the study in the given time or part of the data was corrupted or missing. As a result, the process reported in this section, covers the remaining 18 subjects.

### 3.2.2 Eight Puzzle

According to Johnson and Storey (1879), there are some cases in which the initial 8-puzzle configuration cannot be converted to a goal configuration. For the purpose of this study, four 8-puzzle initial configurations were selected, which could be transformed into their respective goal configuration. The Figure 3.2 shows the configurations. Two puzzles (Figure 3.2, a & b) provide low cost configuration, which have easier goal configuration to remember, and two puzzle configurations (Figure 3.2, c & d) are including high cost of remembering the goal configuration.

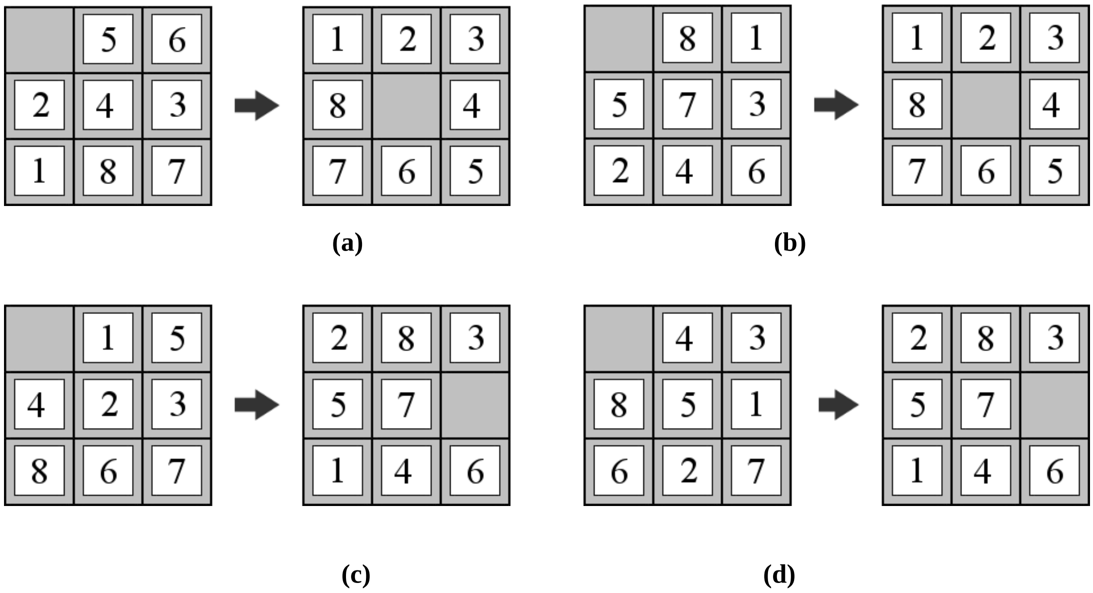



**Figure 3.2. The chosen 8-puzzle configurations for the experiment; each initial configurations is followed by its goal configurations**

### 3.2.3 Apparatus

A version of the 8-puzzle game was implemented in Visual Studio 2008. The 8 puzzle game used point-and-click interaction to select and activate tiles. Selected tiles would then slide into the empty cell. Each tile took exactly 500ms to slide into the empty cell. The puzzle interface was viewed on 23inch monitors, with a screen resolution of 1280 x 1024. At this resolution, each screen button was 200 x 200 pixels. The 8-puzzle software also automatically created a separate log file for each puzzle that recorded button selection data.

In addition to the 8-puzzle log files, eye gaze data for each participant was recorded using Tobii TX300 (300Hz) eye tracker, with at 23’’ screen. Eye movement data were recorded and analyzed using Tobii Studio 3.

Participants sat at a viewing distance of 60cm from the interface and interacted with the puzzle with one of the three interaction methods as follows:

- Gaze-augmented interaction: Eye tracking version of the game is an implementation from Bednarik et al. (2009) study on problem solving enhancement by gaze interaction. The Tobii TX300 eyetracker with the Eye Control Suite 2.1 was used to control the eyes during gaze-augmented interaction. A standard Dell keyboard was also used for Gaze-augmented input. Selection of a tile was achieved by looking at a tile (which then became highlighted), and pressing the spacebar on the keyboard to confirm the selection. Figure 3.3 shows the interaction condition.

**Figure 3.3. The gaze-augmented interaction condition**

- Touch interaction: A HP 2310ti 23" LED touch screen monitor was used for touch interaction. Subject could select tiles on the screen simply by directly pressing on the tile they wanted to select. In order to avoid the user’s hands from interfering with the eye tracker’s line of vision, the eye tracking unit from the TX300 was detached from the eye tracker and fixed on top of the touch screen and inverted. The inverted (flipped upside-down) view of the eye tracker also required that the view of the desktop be inverted (rotated 180 degrees) for accurate eye tracking. Figure 3.4 shows the touch interaction condition.

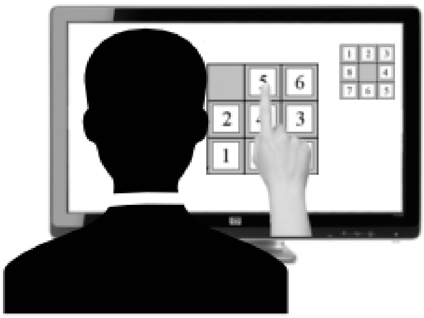


**Figure 3.4. The touch interaction condition**

- Simulated speech interaction: In this condition the user sat in front of the TX300 screen wearing a headset. No mouse or keyboard was present. The user said the number they wished to select into a microphone and thought that the computer was carrying out the selection when in fact a remote human, operator in an adjacent room, was listening to user`s voice commands and selecting the corresponding numbers on the keyboard. The remote operator used TeamViewer 8 software to control the user's computer. Figure 3.5 shows the simulated speech interaction condition.

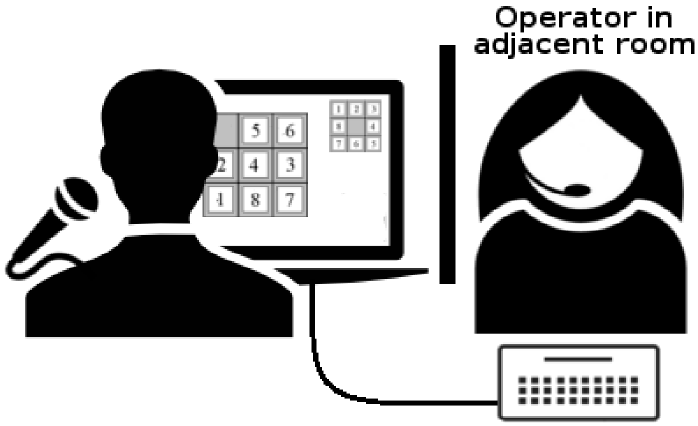



**Figure 3.5. The simulated speech interaction condition**

### 3.2.4 Design and Procedure

The study was conduction in a usability laboratory at the School of Computing, University of Eastern Finland. The usability lab consists of an observation room, with is a one-way mirror wall and a test room.

Participants were first briefed on the particulars of the study and were asked to fill out a pre-test questionnaire and study consent form. Participants were also told that they had exactly 2 hours to complete 6 puzzles, as well as training.

Participants were also informed that they would be left alone in the testing room, and an operator would monitor their progress from the observation room. In order to insure that participants cannot skip any of the tasks, the operator was in charge of navigating the user's views. Once a user completed a task, they informed the operator using their headset that they were ready to move on to the next view.

Each participant was randomly assigned to one of the 3 interaction conditions in the experiment (10 participants for Touch interaction, 6 participants for Gaze-augmented interaction, and 8 participants for Simulated speech interaction), which at the end of the experiment resulted in 6 participants for each interaction method. After assigning participants to interaction methods, participants were seated in front of the testing unit (the screen with the eye tracker and headset), where they first had to pass an eye tracking calibration before starting a series of training tasks. The training tasks were designed to get participants accustomed to the interaction method they were given to solve the task, the user interface and the remote operated view navigation. The training tasks ranged from selecting 1-3 buttons on a screen and solving smaller sliding puzzles, such as the 3-puzzle and 5-puzzle.

Once participants completed training, they could start the experiment. The experiment section included completing six 8-puzzles which were followed by an on-screen NASA Task Load Index questionnaire. The NASA Task Load Index questionnaire is a mental workload evaluation tool which is used to gather the subjective experience of workers engaged in human-machine complex socio-technical systems (Colligan et al., 2015). All participants started with Puzzle (a), see Figure 3.2, and the order of the remaining 3 puzzles were randomized in order to reduce the possible correlation effect of puzzle order on the experiment. Puzzles (b) and (d) were repeated twice, back-to-back.

Once all 6 puzzles were completed, participants were asked to complete a post-test questionnaire to gain further insight on their experiences with their assigned interaction method.

For each 8-puzzle solving task, the inter-move latency, number of tile moves, and total time to solution were recorded inside the 8-puzzle log files. In order to decrease the effect of the remote operator action delay in executing the actions subject asks, the latency between the subject speech command and remote operator action in the first 25 tile moves of the first subject of voice command implementation`s first 8-puzzle were analyzed and the mean latency value was 891 milliseconds (see Appendix 1). Next, the mean value was decreased from the inter-move latencies of the all voice command implementation recorded inter-move latency data. In the case of negative inter-move latency values, the value was replaced with the lowest inter-move latency value of that subject in the corresponding solved 8-puzzle. The detailed description on quantitative measurements and their analysis is provided in Chapter 4.

# 4 Results and discussion

This chapter is divided into two sections; the first section contains a preliminary analysis of recorded experiment data during 8-puzzle problem solving and the differences of the inter-move latencies among the different manipulations, and the second part includes Gaussian mixture model analysis of the inter-move latencies which is to investigate the differences in planning and action phases in problem solving activity using each of the manipulations (In all parts of the analysis, the *level of significance* is set at 0.01).

For the simple naming of the interaction methods, the names Touch, Gaze, and Speech are used for Touch**,** Gaze-augmented and Simulated speech interaction methods, respectively. In addition, the terms ‘subject’, ‘participant’ and ‘user’ are interchangeably used in this chapter and also the following chapters, sections and appendix.

## 4.1 Preliminary analysis

We analyze the available 8-puzzle problem solving process data by considering the solution speed of the subject using different parameters. The data parameters were selected according to Ericsson (1974a) study on relationship of problem solving performance parameters with subjects` 8-puzzle solving. The parameters included total time to solution, inter-move latency, and total number of moves.

A custom MATLAB software script was used to analyze the available data parameters from the raw data inside the recorded log files. MATLAB is high-level language and interactive environment utilized for development of algorithms, visualizing data, data analysis, and numerical computation (Matlab, n.d.).

This section consists of four parts; three parts each contains the analysis of one of the parameters, and one part assigned to discussion of parameters` analysis. The aim is to test the adequacy of the each parameter for subject performance assessment on each of the three used manipulations (The available data for each parameter is given in Appendix 2).

### 4.1.1 Total time to solution

Total time to solution is calculated as follows:

$$\sum_{i=1}^{nj} ti$$

where $t_i$ is the i-th inter-move latency value in the j-th inter-move latency dataset of each interaction method, and $n_j$ is the total number of the inter-move latencies in the j-th inter-move latency dataset of each interaction method. Table 4.1 and Figure 4.1 show the total time to solution analysis of all participants in each interaction method over all puzzles.

**Table 4.1. Mean and SD of total time to solution per interaction method in seconds**

| | Gaze (GA) | Speech (S) | Touch (T) |
|---|---|---|---|
| **Mean** | 269.020 | 274.205 | 176.527 |
| **Standard deviation (SD)** | 214.967 | 340.240 | 123.265 |

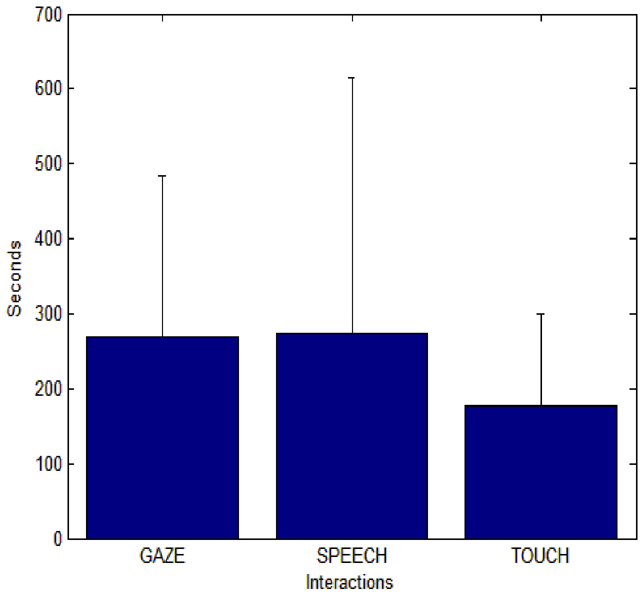



**Figure 4.1. Total time to solution per interaction method**

The observed mean value and standard deviation for Touch interaction method can be interpreted as the lowest time to solution time a participant has used to reach the

solution of the 8-puzzle. For the other interaction methods both mean value and standard deviation were higher than Touch interaction method.
By using the total solution time of participants for each puzzle, the Bartlett's test of homogeneity of variances resulted a significant difference [$\chi^2(2)$ = 32.14, p<0.01].

After log-transformation total solution times dataset was tested again for homogeneity test and Bartlett's test did not show a violation of homogeneity of variances [$\chi^2(2)$ = 1.91, p = 0.38]. With one-way ANOVA, we did not find any significant effects of interaction methods on Total solution time [F(2,105)=2.086, p>0.01]**.** Table 4.2 shows the result.

**Table 4.2. Total solution time of participants in each interaction method ANOVA result**

| | DF | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| **Method** | 2 | 2.69 | 1.343 | 2.086 | 0.129 |
| **Residuals** | 105 | 67.62 | 0.644 | | |

### 4.1.2 Mean inter-move latency

Mean inter-move latency is calculated is follows:

$$\frac{\sum_{i=1}^{nj} ti}{nj}$$

where $t_i$ is the inter-move latency value in the j-th inter-move latency dataset of each interaction method, and $n_j$ is the total number of the inter-move latencies in j-th latency dataset of each interaction method. Figure 4.2 and Table 4.3 show the mean inter-move latency analysis of all participant data in each interaction method over all puzzles.

**Table 4.3. Mean and SD of mean inter-move latency per interaction method in seconds**

| | Gaze (GA) | Speech (S) | Touch (T) |
|---|---|---|---|
| **Mean** | 2.112 | 3.741 | 2.209 |
| **Standard deviation (SD)** | 0.610 | 0.917 | 1.046 |

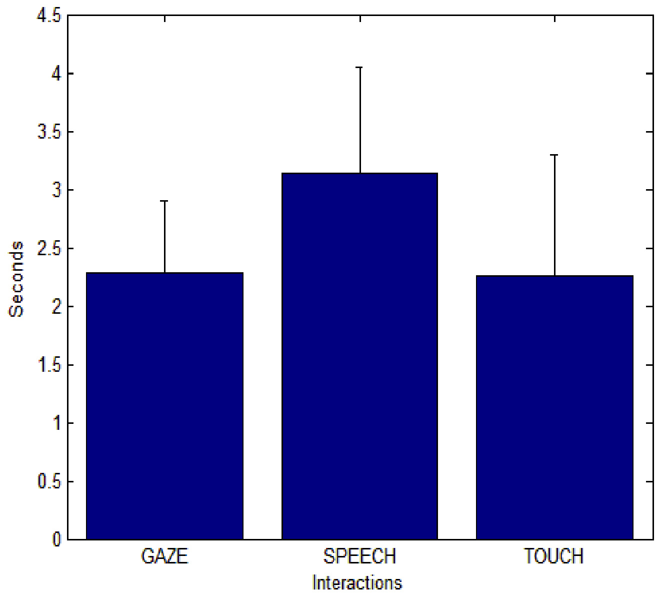



**Figure 4.2. Mean inter-move latency per interaction method**

The observed mean value for Speech interaction method showed a difference with the other two interaction methods. Gaze and Touch interaction method had almost equal mean values but have different standard deviation values.

By using the mean latency of participants for each puzzle the Bartlett's test of homogeneity of variances resulted a significant difference [$\chi^2(2) = 9.79$, $p<0.01$].

After log-transformation, mean latency of participants dataset was tested again for homogeneity test and Bartlett's test did not show a violation of homogeneity of variances [$\chi^2(2) = 4.69$, $p = 0.09$]. With one-way ANOVA, we found a significant effect of Interaction methods on mean latency [$F(2,105)= 14.36$, $p<0.01$]. Table 4.4 shows the result.

**Table 4.4. Mean inter-move latency of participants in each interaction method ANOVA result**

| | **DF** | **Sum Sq** | **Mean Sq** | **F value** | **Pr(>F)** |
|---|---|---|---|---|---|
| **Method** | 2 | 2.812 | 1.4060 | 14.36 | <0.05 |
| **Residuals** | 105 | 10.279 | 0.0979 | | |

For the purpose of the post-hoc test the pairwise t-test with Bonferroni p-value adjustment method is performed and the results are shown in Table 4.5.

**Table 4.5. Bonferroni adjusted p-values pairwise t-test result**

| | Gaze | Speech |
|---|---|---|
| **Speech** | <0.01 | - |
| **Touch** | 1.00 | <0.01 |

From the results in Table 4.5, the same interpretation as Figure 4.2 can be made, and Touch and Gaze interaction methods did not show a significant difference.

### 4.1.3 Total number of moves

Total number of moves per interaction method can be obtained from the available data. Table 4.6 and Figure 4.3 show the total time to solution analysis of all participant data in each interaction method over all puzzles.

**Table 4.6. Mean and SD total number of moves per interaction method**

| | Gaze (GA) | Speech (S) | Touch (T) |
|---|---|---|---|
| **Mean** | 120.084 | 79.944 | 79.222 |
| **Standard deviation (SD)** | 103.339 | 80.217 | 45.317 |

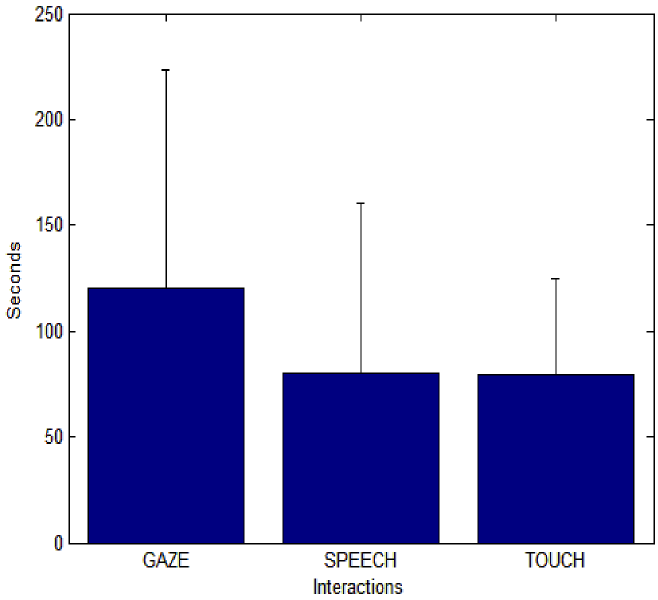



**Figure 4.3. Total number of moves per interaction method**

By using the total moves of participants in each interaction method the Bartlett's test of homogeneity of variances resulted a significant difference [$\chi^2(2) =$ 21.22726, p<0.01].

After log-transformation total solution times dataset was tested again for homogeneity test and Bartlett's test did not show a violation of homogeneity of variances **[**$\chi^2(2)$ **=** 1.89, p = 0.39]. With one-way ANOVA, we did not find any significant effects of interaction method**s** on total number of moves [F(2,105)= 4.25, p>0.01]. Table 4.7 shows the result.

**Table 4.7. Total moves of participants in each interaction method ANOVA result**

| | DF | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| **Method** | 2 | 4.18 | 2.0919 | 4.253 | >0.01 |
| **Residuals** | 105 | 51.65 | 0.4919 | | |

### 4.1.4 Discussion

In the previous subsections, three different parameters were statistically tested to find the possible significant effect of interaction methods on each of them. Total time to solution shown no significant effects of interaction methods, which was caused by the high standard deviation of speech in comparison with the two other interaction methods. In general, Touch interaction methods had the lowest mean (176.527 seconds) and standard deviation (123.265 seconds) compared with Speech and Gaze interaction methods. Similar to Svendson (1991), time to solution for the command mode was quite higher than the direct manipulation mode**.**

The interaction methods caused a significant effect on the mean inter-move latency parameter. By further analysis, Speech interaction method showed a significant difference with the other interaction methods. It was mainly caused by the high difference of mean value for Speech interaction method (3.741 seconds) in compare with Touch and Gaze interaction methods, which had almost the same mean value (2.209 and 2.112 seconds, respectively). The result was alike to previous studies results (O’Hara & Payne, 1998; O’Hara & Payne, 1999) which indicates the higher mental activity for the users of command mode, who had a higher mean inter-move latency.

The total number of the moves showed no significant effects of interaction methods, which cause by higher standard deviation of Gaze interaction method (103.339) compared with the two other interaction methods. The same with the total time to solution parameter, Touch interaction method had the lowest mean (79.222) and standard deviation (45.317) compared with Speech and Gaze interaction method**s.**

In general, these results provided a support for the first hypothesis that inter-move latencies can indicate the performance of subjects during the problem solving, which agrees with the result of Ericsson (1974a) on problem solving performance. Similarly, the current research did not show any effect of total number of tile moves and total time to solution for user problem solving performance assessment.

Moreover, the results provided support for the second hypothesis that the degree of directness of interaction method affects inter-move latency values. As it is discussed above, the users using an interaction method providing a high level of directness (Touch interaction method) had lower mean inter-move latency value in compare with the users on interaction methods proving providing a low level of directness (Gaze and Speech interaction methods) .

## 4.2 Analysis of the inter-move latencies distribution

Following Ericsson`s research (1974b) on latencies distribution, the distribution analysis was utilized to categorize the phases of problem solving into action and planning, and also investigate further about inter-move latencies inside all methods of interaction. In this section, analysis steps have been performed as follows:

1) Log-normal distribution parameter estimation
2) Normalized log-transformed latencies of data distribution
3) Bimodal log-normal distribution parameter estimation

In this section R software environment was used for statistical computing. R is a free software analysis program, which contains a programming language to do a variety of statistical computations and is extended by many open-source developed packages (R project, n.d.).

### 4.2.1 Log-normal distribution parameter estimation

Ericsson's (1974b) research on distribution of inter-move latency data was performed by testing the data distribution similarity to log-normal distribution. The frequency function of the theoretical two-parameter log-normal is given as below (Crow & Shimizu, 1988, p. 2):

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma x}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

A positive value random variable X can have two-parameter lognormal distribution $\Lambda(\mu,\sigma^2)$ if Y=ln X is normally distributed with $N(\mu,\sigma^2)$. As an example of inter-move latency data distribution a histogram of values for the first participant of Gaze manipulation on Puzzle 1 is shown in Figure 4.4.

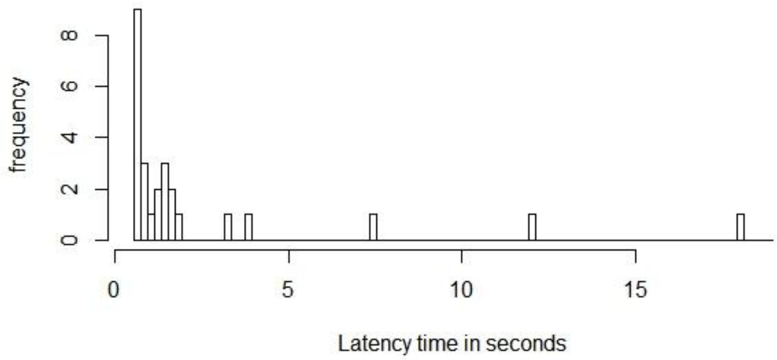



**Figure 4.4. Example of inter-move latency data distribution**

Inspection on the observed inter-move data shows that for many of participant`s puzzle there is no latency data in almost the first 0.5 milliseconds of distributions. To fix the distance from the empirical two-parameter lognormal distribution, the three-parameter lognormal distribution $\Lambda(r,\mu,\sigma^2)$ has been used, where r is the distance of the participant`s puzzle inter-move latency data distribution from the theoretical distribution. It is said that a tree-parameter lognormal has lognormal distribution if Y=log(X-r) is normal distribution with parameters $\mu$ and $\sigma^2$ (Crow & Shimizu, 1988).

For estimating the distance value for each participant`s puzzle inter-move latency data, the eqlnorm3 function from the EnvStat package (see Millard, 2013) in R software environment is used. This function implements different methods for the estimation of three-parameter log normal distribution, and one of the major methods is the Modified Method of Moment Estimators (MMME). Cohen and Whitten (1980) considered MMME as more efficient and accurate than the other methods (For a detailed explaination please refer to Cohen and Whitten, 1980).

After obtaining *r* as a distance parameter, which is mentioned as $A_{Ln}$, for each participant`s inter-move latencies, the distance value is being decreased from each latency value of that participant by the following formula:

$$t'_{ij} = t_{ij} - A_{Ln}$$ , **i:** participant index **j:** inter-move latency data index

After calculating the values of $t'_{ij}$ for each value of each participant`s puzzle inter-move distribution, the standard deviation and mean were calculated. Then, each participant's puzzle fixed distance inter-move latency vector was log-transformed to be assessed for normality using the Pearson chi-square test (The results are provided in Appendix 3).

Most of the subject inter-moves latency vectors agree with the log-normal distribution, and only a few had significant result in the test. After investigation of the histograms of the log-transformed inter-move latency vectors, it was obvious that many of the failed distributions were skewed. In general, 72 out of 108 inter-move latency vectors have had nearly normal distribution, which means nearly 70 percent of the inter-move latency vectors were not violating the normal distribution conditions. On this basis, we consider that majority of them have log-normal distribution and we continue with this assumption to the next section.

### 4.2.2 Normalized log-transformed latencies of data distribution

After fixing the distance for each participant`s puzzle inter-move latency data, for further distribution analysis, due to low amount of data in inter-moves latency data vectors, we decided to combine the distance inter-moves latency data of each puzzle in each interaction method. Then, each of the combined inter-moves latency data vector was transformed to the z-scores, according to Ericsson (1972b), by the formula as follows:

$$t''_{ij} = \frac{\log t'_{ij} - m_{ij}}{s_{ij}}$$ , **i:** inter-move latency data index, **j:** interaction method index

where $m_{ij}$ and $s_{ij}$ are the mean and standard deviation of their corresponding inter-move latency data. We observed that many of the latency data distributions contained at least two modes. By considering the lower mode value as the representative of action distribution and the higher mode as the representative of planning distribution, in the next section we will try to explore the Gaussian mixture model in the inter-move latency data (The histograms of z-scores are available in Appendix 4).

### 4.2.3 Bimodal log-normal distribution parameter estimation

In this section we attempted to investigate the possibility of dividing inter-move latencies into two groups of action and planning clusters. By considering two main modes for each inter-move latency vector distribution, each one as representive of action and planning latency distributions, are considered. In the first step, all fixed distance inter-move latency values are transformed by the following formula in order to transform log-normal data into normal distribution:

$$t^*_i = \log t'_i$$

The next step is to estimate the parameters of the following bimodal normal distribution:

$$f(T) = p\, g_1(t) - (1-p) g_2(t)$$

$$= p\left(\frac{1}{S_1\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{t-m_1}{s_1}\right)}\right) + (1-p)\left(\frac{1}{S_2\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{t-m_2}{s_2}\right)}\right)$$

where $s_1$ and $s_2$ are standard deviation values of each distribution, in addition, $m_1$ and $m_2$ are mean values of each distribution. The bimodal log-normal distribution can be achieved by the insertion of transformed *t* formula in the normal bimodal normal distribution.

For estimating the bimodal normal distribution parameters, the normalMixEm function from mixtools package (Benaglia et al., 2009) in R software environment was utilized. This function utilizes the EM algorithm which tries to maximize the conditional expected complete-data log-likelihood by each M-step of the algorithm execution. The unmixing process of normal distributions using the normalMixEm function is repeated with different initial values in order to maximize the log-likelihood of the estimated distributions. The result of the estimation has been shown in Table 4.8.

**Table 4.8. Estimated parameters for bimodal log-normal distributions in log values**

| Puzzle | Mean 1 | SD 1 | Prop 1 | Mean 2 | SD 2 | Prop 2 |
|---|---|---|---|---|---|---|
| Touch Puzzle 1 | 0.297 | 0.504 | 0.457 | 0.483 | 0.972 | 0.543 |
| Touch Puzzle 2A | -0.09 | 0.991 | 0.934 | 1.978 | 0.868 | 0.066 |
| Touch Puzzle 2B | -1.644 | 0.605 | 0.198 | 0.228 | 1.028 | 0.802 |
| Touch Puzzle 3 | -1.861 | 0.009 | 0.014 | -0.046 | 1.33 | 0.987 |
| Touch Puzzle 4A | -0.613 | 0.107 | 0.083 | 0.233 | 0.959 | 0.917 |
| Touch Puzzle 4B | -1.067 | 0.07 | 0.019 | 0.154 | 1.149 | 0.981 |
| Gaze Puzzle 1 | 0.313 | 0.439 | 0.36 | 0.43 | 0.989 | 0.64 |
| Gaze Puzzle 2A | -1.258 | 0.545 | 0.192 | 0.317 | 1.018 | 0.808 |
| Gaze Puzzle 2B | -1.652 | 0.579 | 0.187 | 0.216 | 1.018 | 0.813 |
| Gaze Puzzle 3 | -1.02 | 0.985 | 0.519 | 0.813 | 0.899 | 0.488 |
| Gaze Puzzle 4A | -0.621 | 0.105 | 0.076 | 0.158 | 1.02 | 0.924 |
| Gaze Puzzle 4B | -1.112 | 0.671 | 0.403 | 0.813 | 0.827 | 0.597 |
| Speech Puzzle 1 | 0.293 | 0.436 | 0.343 | 0.481 | 0.938 | 0.657 |
| Speech Puzzle 2A | -1.217 | 0.447 | 0.127 | 0.267 | 1.099 | 0.874 |
| Speech Puzzle 2B | -1.584 | 0.599 | 0.238 | 0.31 | 1.001 | 0.762 |
| Speech Puzzle 3 | -1.861 | 0.009 | 0.013 | -0.011 | 1.315 | 0.987 |
| Speech Puzzle 4A | -0.622 | 0.11 | 0.099 | 0.215 | 1.021 | 0.901 |
| Speech Puzzle 4B | -0.955 | 0.719 | 0.405 | 0.812 | 0.877 | 0.595 |

Afterwards, the threshold between planning and action distributions (aka components) is investigated manually and it is mentioned as ground truth for testing two other clustering methods (The manually determined thresholds are available in Appendix 5).

In order to be able to automatically do the same threshold determining, two algorithms have been tested on the log-transformed data, hard k-means and a customized version of soft k-means.

The hard k-means is the usual k-means algorithm. It was formulated by Steinhaus in 1956 and was later introduced as k-means by McQueen in 1967 (Kondo et al., 2012). This algorithm is used for clustering data with M points in N dimensions into k (<=n) sets $S=\{s_1,s_2,\ldots,s_k\}$. At first, k centroids are selected by a specific policy (e.g. randomly), later two steps of the algorithm are performed: assigning points to the closest centroid and selecting new centroids by minimizing the within-cluster sum of squares as follows (MacKay, 2003):

$$argmin_s \sum_{i=1}^{k} \sum_{x \in S_i} |x - \mu_i|$$

where $\boldsymbol{\mu}_i$ is the mean of points in $S_i$. This method utilizes the k-means implementation in default stats package of R software environment.

For the soft k-means method, each cluster is described by a Gaussian density. In this algorithm a value of responsibility is computed for each data point and according to the highest responsibility value of a density, the data point is assigned to that density (Hastie et al. 2008). The formulas for a data point responsibility values in a bimodal mixture are:

$$R_1 = g_1(x)/(g_1(x) + g_2(x))$$

$$R_2 = g_2(x)/(g_1(x) + g_2(x))$$

where $g_1$ and $g_2$ are the two existing Gaussian distributions. We customized the soft k-means method for finding the threshold between action and planning distributions. The customized algorithm is shown in Figure 4.5.

1- Sort the inter-move latencies values in the ascending order
2- Set i=0
3- Set i=i+1
4- Compute $A=g_1(t_i)/(g_1(t_i)+g_2(t_i))$ and $B=g_2(t_i)/(g_1(t_i)+g_2(t_i))$ for each inter-move latencies value
5- While A<B repeat Step 3
6- Set i=i+1
7- Compute $A=g_1(t_i)/(g_1(t_i)+g_2(t_i))$ and $B=g_2(t_i)/(g_1(t_i) + g_2(t_i))$ for each inter-move latencies value
8- While A>=B repeat Step 6
9- Compute threshold = $(t_i + t_{i+1})/2$

**Figure 4.5. Modified soft k-means method**

After obtaining the clusters with both algorithms, the mean value of right most value of left cluster and the left-most value of the right cluster, is computed as threshold. For Table 4.9 shows each algorithm obtained threshold and their accuracy considering the ground truth.

The accuracy is computed by the following formula:

$$acc = \frac{\text{number of truely inter} - \text{move latencies assignment to each cluster}}{\text{number of all inter} - \text{move lantecies}}$$

The obtained mean accuracy for hard k-means and modified soft k-means is 0.82 and 0.90, respectively. The modified soft k-means seems to be more close to the ground truth than the hard k-means which is quite sensitive to the extreme inter-move latency values as outliers. For more investigation on the differences between the problem solving phases in each interaction method, the ground truth thresholds (manual thresholds) were transformed into seconds by the following formula:

$$\text{seconds(PT)} = \mathrm{e}^{\text{PT+MD}}$$

where MD (Mean Distance) is the reduced distance value of each puzzle solving activity (for more details refer to section 5.2.1), and PT (Puzzle Threshold) is the manual threshold value of the puzzle. As the inter-move latency values for the same puzzles in each interaction method are combined into one inter-move latency vector, the reduced distance parameter of all combined inter-move latency vectors in each puzzle is averaged and added to the threshold value of each puzzle in each interaction method. Table 4.10 and Figure 4.6 show the results on the obtained transformed values.

**Table 4.9. Threshold and accuracy of manual and algorithmic methods**

| Puzzle | H-K | MS-K | Manual | H-K acc | MS-K acc |
|---|---|---|---|---|---|
| Touch Puzzle 1 | 0.461 | 0.918 | 0.300 | 0.91 | 0.70 |
| Touch Puzzle 2A | 0.088 | 0.953 | 1.000 | 0.71 | 0.98 |
| Touch Puzzle 2B | -0.269 | -0.784 | -1.000 | 0.83 | 0.89 |
| Touch Puzzle 3 | -0.048 | -1.826 | -1.500 | 0.64 | 0.95 |
| Touch Puzzle 4A | 0.302 | -0.384 | -0.200 | 0.78 | 0.94 |
| Touch Puzzle 4B | 0.197 | -0.898 | -0.500 | 0.77 | 0.89 |
| Gaze Puzzle 1 | 0.439 | 0.918 | 0.000 | 0.78 | 0.55 |
| Gaze Puzzle 2A | 0.028 | -0.511 | -0.500 | 0.83 | 1.00 |
| Gaze Puzzle 2B | -0.211 | -0.808 | -1.000 | 0.80 | 0.96 |
| Gaze Puzzle 3 | -0.119 | -0.100 | 0.000 | 0.97 | 0.98 |
| Gaze Puzzle 4A | 0.137 | -0.381 | -0.200 | 0.88 | 0.95 |
| Gaze Puzzle 4B | -0.005 | -0.189 | -0.250 | 0.94 | 0.98 |
| Speech Puzzle 1 | 0.481 | 0.858 | 0.200 | 0.82 | 0.63 |
| Speech Puzzle 2A | 0.061 | -0.530 | -0.500 | 0.84 | 0.99 |
| Speech Puzzle 2B | -0.259 | -0.717 | -0.800 | 0.87 | 0.98 |
| Speech Puzzle 3 | -0.048 | -1.826 | -1.500 | 0.65 | 0.96 |
| Speech Puzzle 4A | 0.235 | -0.381 | -0.200 | 0.84 | 0.95 |
| Speech Puzzle 4B | 0.078 | -0.087 | -0.300 | 0.89 | 0.94 |

**Table 4.10. Mean and SD of mean threshold per interaction method in seconds**

| | Gaze (GA) | Speech (S) | Touch (T) |
|---|---|---|---|
| **Mean** | 1.126 | 0.930 | 1.610 |
| **Standard deviation (SD)** | 0.354 | 0.460 | 1.630 |

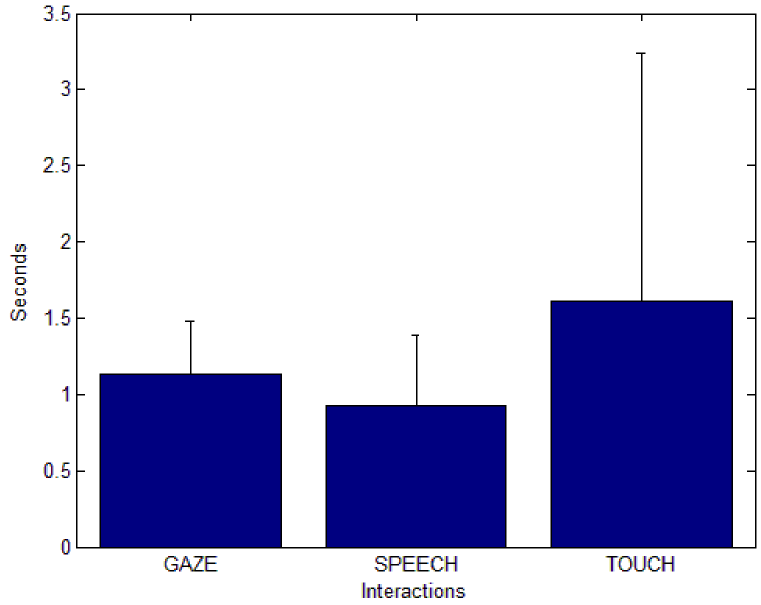



**Figure 4.6. Thresholds of each interaction method in seconds**

### 4.2.4 Discussion

Ericsson (1974b) used the Gaussian mixture analysis to indicate that users can reduce the mental effort and have less planning efforts through learning in the problem solving process. The same analysis process was performed to test the effect of directness on the planfulness of problem solving.

The mean threshold value for Touch interaction method (mean=1.610) was much higher than Gaze and Speech interaction methods value (1.126 and 0.930 seconds, respectively). It can be interpreted that the higher threshold values for the Touch interaction method is a result of more physical actions for users, compared with users who were using the other two interaction methods. However the standard deviation in Touch is the highest which shows that users were more affected by interaction method to determine their planning and action amount, which seems in Gaze and Speech has been the lowest, and seems their users had less effect by interaction method and they were doing almost the planning and action within a short range of difference in each puzzle solving activity.

The higher mean threshold value for Touch interaction method highly agrees with the study of O'Hara (1998, 1999), and it shows that users for the low cost manipulation are more oriented to do more actions on external display (display-based strategy) than the users with higher manipulation cost, who were using more planning based strategy. The users who were using the Gaze and Speech interaction methods had closer mean thresholds.

The Touch interaction method provided similar findings to Kieras et al. (2001), whereby it provides a direct manipulation capability and also a narrow gulf of execution. Gaze and Speech interaction shown users were more oriented to a higher amount of planning, which was in agreement with the results of Bednarik et al. (2009) and discussion by Frohlich (1997).

# 5 Conclusion

In this research work, the problem solving activity with different interaction methods was studied with the aim of revealing the effect of interaction method directness level on strategy of user problem solving. The utilized interaction methods were selected according to their directness level, Touch interaction method which corresponds to direct manipulation mode, as well as Gaze-augmented and Simulated speech interaction methods which corresponds to command mode. The effect of interaction method directness level on strategy of user problem solving was assessed by Gaussian mixture analysis. In addition to this analysis, different quantitative measurements were tested to check if they can be good enough for the user performance study.

The ANOVA test was applied on the quantitative measurements and it provided a support for two hypothesizes that the inter-move latencies can be used to measure problem solving performance for the 8-tile puzzle game, and the amount of interaction method directness affect inter-move latency. A significant effect of Interactions on mean latency [$F(2,105)= 14.36$, $p<0.01$] was found and the touch interaction method as well as gaze-augmented interaction method did not show a significant difference.

The touch interaction method has shown to be more oriented to direct manipulation mode, by having both low mean total time to solution and low mean inter-move latency. Whereas, the Simulated speech interaction was the orientated to command-mode manipulation mode with high mean inter-move latencies and high total time to solution. The Gaze-augmented interaction did not give a clear position for justifying its manipulation modes by the initial results.

An exploration of Gaussian mixture models on the inter-move latency data, supported the hypothesis that inter-move latency of 8-puzzle can have effects on problem solving processes. According to the initial result, the inter-move latency was affected by the directness degree of the interactions. A further analysis of the effect of this behavior on the user problem solving has shown that there are at least two log-normal distributions included in each inter-move latencies vector.

The final phase of research, on the categorizing of latency distributions into action and planning resulted in a higher proportion of the actions for the touch interaction method (mean=1.610, SD=1.630), and in a lower proportion of actions for Gaze-augmented (mean=1.126, SD=0.354) and Simulated speech (mean=0.930, SD=0.460) interaction methods. It was a clear sign of users' display-based strategy for solving the 8-puzzle. Generally, the results of research show users with the highest amount of directness (Touch interaction method) were able to off-load their mental efforts into the external representation, and users with less level of directness (Gaze-augmented and Simulated speech interaction methods) had their load of mental efforts inside their internal representation.

The further study of this research can be analysis of the details involved inside the 8-puzzle tile moves and extending the quantitative parameters, which can help to understand how user decides to change the strategy of solving the problem between off-loading cognitive work to environment and keeping the load of cognitive work inside the working memory.

# References

Adams, B. (2009). *The Psychology Companion*. Palgrave Macmillan.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, *48*(1), 35-44.

Anderson, J. R. (2000). *Cognitive psychology and its implications*. WH Freeman/Times Books/Henry Holt & Co.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation*, 2, 89-195.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory?. *Trends in cognitive sciences*, 4(11), 417-423.

Bednarik, R., Gowases, T., & Tukiainen, M. (2009). Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience. *Journal of Eye Movement Research*, 3(1), 1-10.

Benaglia, T., Chauveau, D., Hunter, D., & Young, D. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6), 1-29.

Byrne, M. D. (2005). Cognitive architectures in HCI: Present work and future directions. In *Proceedings of the 11th International Conference on Human Computer Interaction, Las Vegas, NV*.

Card, S., Moran, T., and Newell, A. (1986). The Model Human Processor. In Boff, K., Kaufman, L., and Thomas, J. (Eds.) Handbook of Perception and Human Performance. Vol. II. Wiley, 45:1-45:35

Cohen, A. C., & Whitten, B. J. (1980). Estimation in the three-parameter lognormal distribution. *Journal of the American Statistical Association*, 75(370), 399-404.

Colligan, L., Potts, H. W., Finn, C. T., & Sinkin, R. A. (2015). Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International journal of medical informatics*, *84*(7), 469-476.

Coon, D., & Mitterer, J. O. (2010). *Introduction to Psychology: Gateways to Mind and Behavior*. Wadsworth Cengage Learning.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin*, 104(2), 163-191.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169, 323-338.

Crow, E. L., & Shimizu, K. (Eds.). (1988). *Lognormal distributions: Theory and applications* (Vol. 88). New York: Dekker.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 3(2), 255-267.

Dunbar, K. (1998). Problem solving. *A companion to cognitive science*, 289-298.

Duncker, K. (1945). On problem-solving. *Psychological monographs*, 58(5).

Ericsson, K. A. (1974a). *Problem-solving behaviour with the 8-puzzle I: time to solution.* Report No. 431, Department of Psychology, University of Stockholm.

Ericsson, K. A. (1974b). *Problem-solving behaviour with the 8-puzzle II: distribution of latencies.* Report No. 432, Department of Psychology, University of Stockholm

Frohlich, D. M. (1997). Direct manipulation and other lessons. *Handbook of human-computer interaction*, 463-488.

Giacoppo, A. S. (2001). The role of theory in HCI. *CHARM: Choosing human–computer interaction (HCI) appropriate research methods.* University of Maryland. http://www. otal. umd. edu/charm.

Goldstein, E. (2008). *Cognitive psychology: Connecting mind, research and everyday experience*. Cengage Learning.

Groome, D. (2013). *An introduction to cognitive psychology: Processes and disorders*. Psychology Press.

Guthrie, *E. R.*(*1952*). *The psychology of learning* (*Rev*. ed.). New York: Harper & Brothers.

Hardin, L. E. (2003). Problem-solving concepts and theories. *Journal of veterinary medical education*, *3*0(3), 226-229.

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics.

Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, 28(3), 249-276.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive science*, 3(4), 275-310.

Heeter C. 1991. The Look and Feel of Direct Manipulation, HYPERNEXUS: *Journal of Hypermedia and Multimedia Studies*, Fall.

Henry, L. (2011). *The development of working memory in children*. Sage.

Hornecker, E. (2005, September). A Design Theme for Tangible Interaction: Embodied Facilitation. In *ECSCW, Vol. 5*, 23-43.

House, E. B., & House, W. J. (1987). Problem-solving: The debates in composition and psychology. *Journal of Advanced Composition*, 62-75.

Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science*.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human–Computer Interaction*, 1(4), 311-338.

Johnson, W. W., & Story, W. E. (1879). Notes on the " 15" puzzle. *American Journal of Mathematics*, 2(4), 397-404.

Kieras, D., Meyer, D., & Ballas, J. (2001, March). Towards demystification of direct manipulation: Cognitive modeling charts the gulf of execution. *Proceedings of the SIGCHI conference on human factors in computing system.* New York, ACM. 128-135

Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18,* 513–549.

Kohler, W. (1927). *The Mentality Of Apes,* Routledge & Kegan Paul, London. 2nd edition.

Kondo, Y., Salibian-Barrera, M., & Zamar, R. (2012). A robust and sparse K-means clustering algorithm. *arXiv preprint arXiv:1201.6082*.

Laurel, B. (1986). Interface as mimesis. *User centered system design: New perspectives on human-computer interaction*, 67-85.

Laurillard, D. (1997). Styles and approaches in problem-solving, in Marton, F., Hounsell, D. and Entwistle, N.J. (eds.), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2nd edition). Edinburgh: Scottish Academic Press, pp. 126–144.

Lovett, M. C. (2002). Problem solving. In D. Medin (Ed.), *Stevens' handbook of experimental psychology: Memory and cognitive processes*. New York. 317–362

Lu, Z. L., & Dosher, B. A. (2007). Cognitive psychology. *Scholarpedia*, 2(8), 2769.

Luine, V. N. (2014). Estradiol and cognitive function: Past, present and future. *Hormones and behavior*, 66(4), 602-618.

MacKay, D. (2003). An example inference task: Clustering. *Information Theory, Inference and Learning Algorithms*, 284-292.

Manches, A., O'Malley, C., & Benford, S. (2009, February). Physical manipulation: evaluating the potential for tangible designs. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction.* ACM. 77-84.

Matlab (n.d.). *Homepage*. Retrieved 2015-04-12 from http://www.mathworks.com/.

McLeod, S. A. (2012). *Working Memory*. Retrieved 2015-05-25 from www.simplypsychology.org/working%20memory.html

Millard, S. P. (2013). Package for Environmental Statistics, including US EPA Guidance.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2), 81.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior.* Holt, Rinehart & Winston, London.

Moran, T. P. (1983, December). Getting into a system: External-internal task mapping analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* . ACM. 45-49.

Neisser, U. (2014). *Cognitive Psychology*: Classic Edition. Psychology Press.

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Newell, A., & Card, S. K. (1985). The prospects for psychological science in human-computer interaction. *Human-computer interaction, 1*(3), 209-242.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1984). Stages and levels in human-machine interaction. *International journal of man-machine studies*, 21(4), 365-375.

Norman, D. A. (1991). Cognitive artifacts. In: *J.M. Carroll* (*ed.*) *Designing Interaction*:*Psychology* at the *Human-computer Interface*. Cambridge: *Cambridge University Press*.

Norman, D. A. (2002). *The design of everyday things: Revised and expanded edition*. Basic books.

Norman, D. A. (1986) Cognitive engineering. In D. A. Norman and S. W. Draper (eds.) *User centered system design*. Hillsdale, NJ: Lawrence Erlbaum Associates, 31-62.

O'Hara, K. (1994). Cost of operations affects planfulness of problem-solving behaviour. In *Conference companion on Human factors in computing systems*. ACM. 105-106.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive psychology*, *35*(1), 34-70.

O'Hara, K. P., & Payne, S. J. (1999). Planning and the user interface: The effects of lockout time and error recovery cost. *International Journal of Human-Computer Studies*, 50(1), 41-59.

Ohlsson, S. (2012). The problems with problem solving: Reflections on the rise, current status, and possible future of a cognitive research paradigm. *The Journal of Problem Solving*, 5(1), 101-128.

R Project (n.d.). *What is R?*. Retrieved 2015-05-30 from http://www.r-project.org/about.html.

Reinefeld, A. (1993). Complete solution of the Eight-Puzzle and the benefit of node-ordering in IDA*. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 248–253.

Richardson-Klavehn, A. and Bjork, R. A. (2002). Memory, Long-Term. *Encyclopedia of Cognitive Science*, 1096-1105

Robertson, S. P., & Black, J. B. (1986). Structure and development of plans in computer text editing. *Human-Computer Interaction*, *2*(3), 201-226.

Schacter, D. L., & Tulving, E. (1994). *Memory systems 1994*. MIT Press. 2-11.

Sedig, K., Klawe, M., & Westrom, M. (2001). Role of interface manipulation style and scaffolding on cognition and concept learning in learnware. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *8*(1), 34-59.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review*, 84(2), 127.

Shih, R., Dubrowski, A., & Carnahan, H. (2009, March). Evidence for haptic memory. In *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint* . IEEE. 145-149

Shneiderman, B. (1992). *Designing the user interface: strategies for effective human-computer interaction* (Vol. 3). Reading, MA: Addison-Wesley.

Shneiderman, B. (1993). 1.1 direct manipulation: A step beyond programming languages. *Sparks of innovation in human-computer interaction*, 17.

Smith, J. B. (1994). *Collective intelligence in computer-based collaboration*. CRC Press.

Sperling, G. (1960). The information available in brief visual presentations.*Psychological monographs: General and applied*, *74*(11), 1.

Sternberg, R. J., & Sternberg, K. (2012). *Cognitive psychology*. Cengage Learning.

Sternberg, R. J., & Ruzgis, P. (1994). *Personality and intelligence*. Cambridge University Press.

Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge university press.

Svendsen, G. B. (1991). The influence of interface style on problem solving.*International Journal of Man-Machine Studies*, 35(3), 379-397.

Thorndike, E. L. (1898). Animal intelligence: an experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4).

Tolman, E. C. (1951). *Purposive behavior in animals and men*. Univ of California Press.

Tulving, E. (1995). Organization of memory: Quo vadis. *The cognitive neurosciences*, *839847*.

Winn, W., & Snyder, D. (1996). Cognitive perspectives in psychology. *Handbook of research for educational communications and technology: A project of the Association for Educational Communications and Technology*, 79-112.

Wray, R. E., & Jones, R. M. (2006). Considering Soar as an agent architecture. *Cognition and multi-agent interaction: From cognitive modeling to social simulation*, 53-78.

Zhang, J. (1991). The interaction of internal and external representations in a problem solving task. In *Proceedings of the thirteenth annual conference of cognitive science society* (Vol. 88, p. 91). Erlbaum Hillsdale, NJ.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive science*, 18(1), 87-122.

# Appendix

**1 Remote operator action delay; delay values in the first 25 tile moves of the first user of Speech interaction method`s first 8-puzzle**

| Tile move index | Remote operator action delay |
|---|---|
| 1 | 0.868 |
| 2 | 0.785 |
| 3 | 0.149 |
| 4 | 0.518 |
| 5 | 0.678 |
| 6 | 1.855 |
| 7 | 1.979 |
| 8 | 0.701 |
| 9 | 0.711 |
| 10 | 0.508 |
| 11 | 1.379 |
| 12 | 0.402 |
| 13 | 0.770 |
| 14 | 1.550 |
| 15 | 0.923 |
| 16 | 1.025 |
| 17 | 0.530 |
| 18 | 0.996 |
| 19 | 0.791 |
| 20 | 0.887 |
| 21 | 0.503 |
| 22 | 1.149 |
| 23 | 0.562 |
| 24 | 0.426 |
| 25 | 1.640 |

**2 Data parameters values; according to the users in each interaction method and 8-puzzle**

| User data Index | Total time to solution | Mean inter-move latency | Total number of moves |
|---|---|---|---|
| User 1 Touch Puzzle 1 | 88.284 | 2.207 | 40 |
| User 1 Touch Puzzle 2A | 50.345 | 1.573 | 32 |
| User 1 Touch Puzzle 2B | 108.287 | 1.547 | 70 |
| User 1 Touch Puzzle 3 | 158.962 | 1.939 | 82 |
| User 1 Touch Puzzle 4A | 70.994 | 1.919 | 37 |
| User 1 Touch Puzzle 4B | 49.368 | 1.646 | 30 |
| User 2 Touch Puzzle 1 | 70.860 | 3.543 | 20 |
| User 2 Touch Puzzle 2A | 462.615 | 3.258 | 142 |
| User 2 Touch Puzzle 2B | 297.512 | 3.42 | 87 |
| User 2 Touch Puzzle 3 | 203.931 | 2.872 | 71 |
| User 2 Touch Puzzle 4A | 349.710 | 4.427 | 79 |
| User 2 Touch Puzzle 4B | 426.466 | 3.986 | 107 |
| User 3 Touch Puzzle 1 | 121.036 | 1.593 | 76 |
| User 3 Touch Puzzle 2A | 82.811 | 1.294 | 64 |
| User 3 Touch Puzzle 2B | 56.285 | 1.309 | 43 |
| User 3 Touch Puzzle 3 | 229.257 | 1.479 | 155 |
| User 3 Touch Puzzle 4A | 39.096 | 1.700 | 23 |
| User 3 Touch Puzzle 4B | 65.762 | 1.399 | 47 |
| User 4 Touch Puzzle 1 | 239.438 | 6.301 | 38 |
| User 4 Touch Puzzle 2A | 149.921 | 2.585 | 58 |
| User 4 Touch Puzzle 2B | 492.516 | 2.706 | 182 |
| User 4 Touch Puzzle 3 | 317.191 | 2.538 | 125 |
| User 4 Touch Puzzle 4A | 231.968 | 2.697 | 86 |
| User 4 Touch Puzzle 4B | 59.942 | 1.934 | 31 |
| User 5 Touch Puzzle 1 | 179.038 | 1.738 | 103 |
| User 5 Touch Puzzle 2A | 155.733 | 1.527 | 102 |
| User 5 Touch Puzzle 2B | 249.309 | 1.578 | 158 |
| User 5 Touch Puzzle 3 | 188.452 | 1.584 | 119 |

| User data Index | Total time to solution | Mean inter-move latency | Total number of moves |
|---|---|---|---|
| User 5 Touch Puzzle 4A | 125.321 | 1.671 | 75 |
| User 5 Touch Puzzle 4B | 340.958 | 1.804 | 189 |
| User 6 Touch Puzzle 1 | 82.447 | 1.874 | 44 |
| User 6 Touch Puzzle 2A | 110.380 | 1.903 | 58 |
| User 6 Touch Puzzle 2B | 97.577 | 1.394 | 70 |
| User 6 Touch Puzzle 3 | 212.631 | 2.025 | 105 |
| User 6 Touch Puzzle 4A | 141.240 | 1.702 | 83 |
| User 6 Touch Puzzle 4B | 49.315 | 2.348 | 21 |
| User 1 Gaze Puzzle 1 | 66.209 | 2.547 | 26 |
| User 1 Gaze Puzzle 2A | 103.807 | 1.922 | 54 |
| User 1 Gaze Puzzle 2B | 270.262 | 2.350 | 115 |
| User 1 Gaze Puzzle 3 | 188.497 | 2.327 | 81 |
| User 1 Gaze Puzzle 4A | 60.598 | 1.638 | 37 |
| User 1 Gaze Puzzle 4B | 48.592 | 1.388 | 35 |
| User 2 Gaze Puzzle 1 | 299.204 | 2.992 | 100 |
| User 2 Gaze Puzzle 2A | 305.063 | 2.630 | 116 |
| User 2 Gaze Puzzle 2B | 187.181 | 2.600 | 72 |
| User 2 Gaze Puzzle 3 | 156.952 | 2.211 | 71 |
| User 2 Gaze Puzzle 4A | 223.708 | 2.762 | 81 |
| User 2 Gaze Puzzle 4B | 108.111 | 2.514 | 43 |
| User 3 Gaze Puzzle 1 | 234.869 | 2.553 | 92 |
| User 3 Gaze Puzzle 2A | 195.073 | 1.876 | 104 |
| User 3 Gaze Puzzle 2B | 410.772 | 1.834 | 224 |
| User 3 Gaze Puzzle 3 | 186.141 | 2.417 | 77 |
| User 3 Gaze Puzzle 4A | 280.046 | 2.394 | 117 |
| User 3 Gaze Puzzle 4B | 240.455 | 2.091 | 115 |
| User 4 Gaze Puzzle 1 | 697.419 | 1.685 | 414 |
| User 4 Gaze Puzzle 2A | 129.959 | 1.688 | 77 |
| User 4 Gaze Puzzle 2B | 414.597 | 1.502 | 276 |

| User data Index | Total time to solution | Mean inter-move latency | Total number of moves |
|---|---|---|---|
| User 4 Gaze Puzzle 3 | 228.006 | 1.689 | 135 |
| User 4 Gaze Puzzle 4A | 461.605 | 1.560 | 296 |
| User 4 Gaze Puzzle 4B | 48.326 | 1.790 | 27 |
| User 5 Gaze Puzzle 1 | 47.268 | 1.477 | 32 |
| User 5 Gaze Puzzle 2A | 44.229 | 1.382 | 32 |
| User 5 Gaze Puzzle 2B | 293.252 | 2.327 | 126 |
| User 5 Gaze Puzzle 3 | 328.056 | 2.504 | 131 |
| User 5 Gaze Puzzle 4A | 60.598 | 1.638 | 37 |
| User 5 Gaze Puzzle 4B | 48.592 | 1.388 | 35 |
| User 6 Gaze Puzzle 1 | 299.204 | 2.992 | 100 |
| User 6 Gaze Puzzle 2A | 305.063 | 2.630 | 116 |
| User 6 Gaze Puzzle 2B | 187.181 | 2.600 | 72 |
| User 6 Gaze Puzzle 3 | 156.952 | 2.211 | 71 |
| User 6 Gaze Puzzle 4A | 223.708 | 2.762 | 81 |
| User 6 Gaze Puzzle 4B | 108.111 | 2.514 | 43 |
| User 1 Speech Puzzle 1 | 211.162 | 2.373 | 89 |
| User 1 Speech Puzzle 2A | 149.460 | 2.231 | 67 |
| User 1 Speech Puzzle 2B | 367.921 | 3.754 | 98 |
| User 1 Speech Puzzle 3 | 381.556 | 3.469 | 110 |
| User 1 Speech Puzzle 4A | 1197.516 | 2.395 | 500 |
| User 1 Speech Puzzle 4B | 310.951 | 3.417 | 91 |
| User 2 Speech Puzzle 1 | 330.413 | 2.924 | 113 |
| User 2 Speech Puzzle 2A | 477.486 | 3.205 | 149 |
| User 2 Speech Puzzle 2B | 93.222 | 2.453 | 38 |
| User 2 Speech Puzzle 3 | 207.159 | 4.143 | 50 |
| User 2 Speech Puzzle 4A | 234.063 | 4.501 | 52 |
| User 2 Speech Puzzle 4B | 242.248 | 4.250 | 57 |
| User 3 Speech Puzzle 1 | 105.671 | 3.409 | 31 |
| User 3 Speech Puzzle 2A | 90.103 | 2.730 | 33 |

| User data Index | Total time to solution | Mean inter-move latency | Total number of moves |
|---|---|---|---|
| User 3 Speech Puzzle 2B | 137.466 | 2.546 | 54 |
| User 3 Speech Puzzle 3 | 255.331 | 1.995 | 128 |
| User 3 Speech Puzzle 4A | 112.717 | 2.087 | 54 |
| User 3 Speech Puzzle 4B | 109.876 | 2.073 | 53 |
| User 4 Speech Puzzle 1 | 227.238 | 2.206 | 103 |
| User 4 Speech Puzzle 2A | 468.006 | 2.871 | 163 |
| User 4 Speech Puzzle 2B | 94.927 | 2.566 | 37 |
| User 4 Speech Puzzle 3 | 201.152 | 2.579 | 78 |
| User 4 Speech Puzzle 4A | 1178.534 | 5.333 | 221 |
| User 4 Speech Puzzle 4B | 284.602 | 3.603 | 79 |
| User 5 Speech Puzzle 1 | 94.852 | 3.162 | 30 |
| User 5 Speech Puzzle 2A | 201.975 | 3.482 | 58 |
| User 5 Speech Puzzle 2B | 195.127 | 2.788 | 70 |
| User 5 Speech Puzzle 3 | 108.74 | 3.508 | 31 |
| User 5 Speech Puzzle 4A | 1623.696 | 3.812 | 426 |
| User 5 Speech Puzzle 4B | 80.533 | 4.239 | 19 |
| User 6 Speech Puzzle 1 | 95.287 | 2.978 | 32 |
| User 6 Speech Puzzle 2A | 65.902 | 1.938 | 34 |
| User 6 Speech Puzzle 2B | 93.152 | 2.025 | 46 |
| User 6 Speech Puzzle 3 | 33.159 | 1.951 | 17 |
| User 6 Speech Puzzle 4A | 61.436 | 2.457 | 25 |
| User 6 Speech Puzzle 4B | 49.638 | 2.256 | 22 |

**3 Estimated parameters for log-normal distributions and Pearson chi-square goodness of fit results**

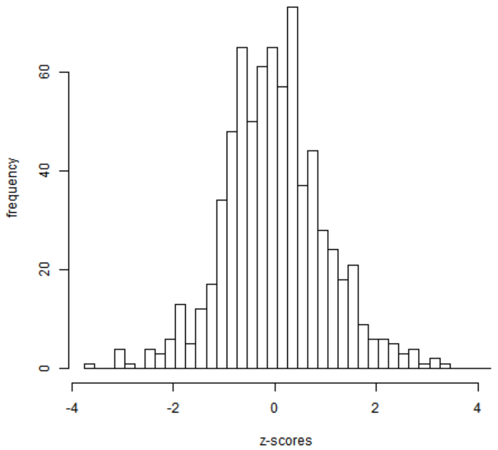
| User data Index | $A_{Ln}$ | m | s | df | $X^2$ | P |
|---|---|---|---|---|---|---|
| User 1 Touch Puzzle 1 | 0.409 | -0.133 | 1.154 | 6 | 26.6 | <0.01 |
| User 1 Touch Puzzle 2A | 0.504 | -0.828 | 1.319 | 5 | 10 | 0.075 |
| User 1 Touch Puzzle 2B | 0.485 | -0.777 | 1.296 | 8 | 22.086 | <0.01 |
| User 1 Touch Puzzle 3 | -0.009 | 0.108 | 0.913 | 9 | 75.756 | <0.01 |
| User 1 Touch Puzzle 4A | 0.465 | -0.745 | 1.418 | 6 | 6.541 | 0.365 |
| User 1 Touch Puzzle 4B | -0.245 | 0.394 | 0.667 | 5 | 19.6 | <0.01 |
| User 2 Touch Puzzle 1 | 0.436 | -0.821 | 1.663 | 4 | 13.6 | <0.01 |
| User 2 Touch Puzzle 2A | 0.409 | -0.139 | 1.613 | 12 | 48.141 | <0.01 |
| User 2 Touch Puzzle 2B | 0.095 | 0.381 | 1.233 | 9 | 41.966 | <0.01 |
| User 2 Touch Puzzle 3 | 0.44 | -0.817 | 1.754 | 9 | 49.845 | <0.01 |
| User 2 Touch Puzzle 4A | 0.395 | -0.24 | 1.847 | 9 | 36.291 | <0.01 |
| User 2 Touch Puzzle 4B | 0.44 | -0.45 | 1.796 | 10 | 33.084 | <0.01 |
| User 3 Touch Puzzle 1 | 0.715 | -0.86 | 1.116 | 9 | 7.053 | 0.632 |
| User 3 Touch Puzzle 2A | 0.801 | -1.115 | 0.878 | 8 | 9.906 | 0.272 |
| User 3 Touch Puzzle 2B | -1.838 | 1.133 | 0.159 | 7 | 31.651 | <0.01 |
| User 3 Touch Puzzle 3 | 0.648 | -0.617 | 0.904 | 13 | 16.665 | 0.215 |
| User 3 Touch Puzzle 4A | 0.717 | -0.533 | 1.009 | 5 | 10.044 | 0.074 |
| User 3 Touch Puzzle 4B | 0.79 | -0.882 | 0.848 | 7 | 11.511 | 0.119 |
| User 4 Touch Puzzle 1 | 0.354 | 0.689 | 1.478 | 6 | 5.579 | 0.472 |
| User 4 Touch Puzzle 2A | 0.501 | -0.002 | 1.258 | 8 | 9.517 | 0.301 |
| User 4 Touch Puzzle 2B | 0.437 | 0.071 | 1.229 | 14 | 24.615 | 0.039 |
| User 4 Touch Puzzle 3 | 0.376 | 0.179 | 1.114 | 11 | 14.888 | 0.188 |
| User 4 Touch Puzzle 4A | 0.334 | 0.311 | 1.016 | 9 | 21.721 | <0.01 |
| User 4 Touch Puzzle 4B | 0.387 | -0.347 | 1.215 | 5 | 13.645 | 0.018 |
| User 5 Touch Puzzle 1 | 0.079 | 0.223 | 0.695 | 10 | 61.961 | <0.01 |
| User 5 Touch Puzzle 2A | 0.525 | -0.627 | 1.053 | 10 | 41.255 | <0.01 |
| User 5 Touch Puzzle 2B | 0.503 | -0.614 | 1.138 | 13 | 54.051 | <0.01 |
| User 5 Touch Puzzle 3 | 0.506 | -0.752 | 1.038 | 11 | 102.059 | <0.01 |

| Participants | $A_{Ln}$ | m | s | df | $X^2$ | P |
|---|---|---|---|---|---|---|
| User 5 Touch Puzzle 4A | 0.409 | -0.133 | 1.154 | 6 | 26.6 | <0.01 |
| User 5 Touch Puzzle 4B | 0.504 | -0.828 | 1.319 | 5 | 10 | 0.075 |
| User 6 Touch Puzzle 1 | 0.485 | -0.777 | 1.296 | 8 | 22.086 | <0.01 |
| User 6 Touch Puzzle 2A | -0.009 | 0.108 | 0.913 | 9 | 75.756 | <0.01 |
| User 6 Touch Puzzle 2B | 0.465 | -0.745 | 1.418 | 6 | 6.541 | 0.365 |
| User 6 Touch Puzzle 3 | -0.245 | 0.394 | 0.667 | 5 | 19.6 | <0.01 |
| User 6 Touch Puzzle 4A | 0.436 | -0.821 | 1.663 | 4 | 13.6 | <0.01 |
| User 6 Touch Puzzle 4B | 0.409 | -0.139 | 1.613 | 12 | 48.141 | <0.01 |
| User 1 Gaze Puzzle 1 | 0.095 | 0.381 | 1.233 | 9 | 41.966 | <0.01 |
| User 1 Gaze Puzzle 2A | 0.44 | -0.817 | 1.754 | 9 | 49.845 | <0.01 |
| User 1 Gaze Puzzle 2B | 0.395 | -0.24 | 1.847 | 9 | 36.291 | <0.01 |
| User 1 Gaze Puzzle 3 | 0.44 | -0.45 | 1.796 | 10 | 33.084 | <0.01 |
| User 1 Gaze Puzzle 4A | 0.715 | -0.86 | 1.116 | 9 | 7.053 | 0.632 |
| User 1 Gaze Puzzle 4B | 0.801 | -1.115 | 0.878 | 8 | 9.906 | 0.272 |
| User 2 Gaze Puzzle 1 | -1.838 | 1.133 | 0.159 | 7 | 31.651 | <0.01 |
| User 2 Gaze Puzzle 2A | 0.648 | -0.617 | 0.904 | 13 | 16.665 | 0.215 |
| User 2 Gaze Puzzle 2B | 0.717 | -0.533 | 1.009 | 5 | 10.044 | 0.074 |
| User 2 Gaze Puzzle 3 | 0.79 | -0.882 | 0.848 | 7 | 11.511 | 0.119 |
| User 2 Gaze Puzzle 4A | 0.354 | 0.689 | 1.478 | 6 | 5.579 | 0.472 |
| User 2 Gaze Puzzle 4B | 0.501 | -0.002 | 1.258 | 8 | 9.517 | 0.301 |
| User 3 Gaze Puzzle 1 | 0.437 | 0.071 | 1.229 | 14 | 24.615 | 0.039 |
| User 3 Gaze Puzzle 2A | 0.437 | 0.071 | 1.229 | 14 | 24.615 | 0.039 |
| User 3 Gaze Puzzle 2B | 0.376 | 0.179 | 1.114 | 11 | 14.888 | 0.188 |
| User 3 Gaze Puzzle 3 | 0.334 | 0.311 | 1.016 | 9 | 21.721 | <0.01 |
| User 3 Gaze Puzzle 4A | 0.387 | -0.347 | 1.215 | 5 | 13.645 | 0.018 |
| User 3 Gaze Puzzle 4B | 0.079 | 0.223 | 0.695 | 10 | 61.961 | <0.01 |
| User 4 Gaze Puzzle 1 | 0.525 | -0.627 | 1.053 | 10 | 41.255 | <0.01 |
| User 4 Gaze Puzzle 2A | 0.503 | -0.614 | 1.138 | 13 | 54.051 | <0.01 |
| User 4 Gaze Puzzle 2B | 0.437 | 0.071 | 1.229 | 14 | 24.615 | 0.039 |

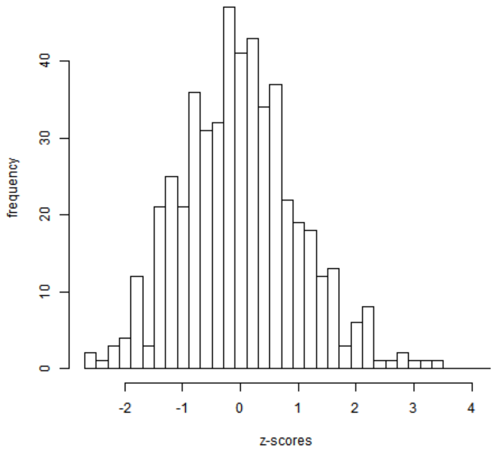
| Participants | $A_{Ln}$ | m | s | df | $X^2$ | P |
|---|---|---|---|---|---|---|
| User 4 Gaze Puzzle 3 | 0.376 | 0.179 | 1.114 | 11 | 14.888 | 0.188 |
| User 4 Gaze Puzzle 4A | 0.075 | 0.093 | 0.734 | 17 | 113.73 | <0.01 |
| User 4 Gaze Puzzle 4B | 0.162 | 0.154 | 0.876 | 5 | 4.704 | 0.453 |
| User 5 Gaze Puzzle 1 | 0.069 | 0.221 | 0.514 | 5 | 5.5 | 0.358 |
| User 5 Gaze Puzzle 2A | 0.496 | -0.823 | 1.215 | 5 | 8 | 0.156 |
| User 5 Gaze Puzzle 2B | 0.433 | -0.331 | 1.417 | 11 | 25.333 | <0.01 |
| User 5 Gaze Puzzle 3 | 0.466 | -0.348 | 1.487 | 12 | 46.137 | <0.01 |
| User 5 Gaze Puzzle 4A | 0.426 | -0.026 | 1.216 | 10 | 9.888 | 0.450 |
| User 5 Gaze Puzzle 4B | 0.454 | -0.287 | 1.345 | 8 | 13.284 | 0.103 |
| User 6 Gaze Puzzle 1 | 0.403 | 0.765 | 0.983 | 10 | 11.306 | 0.334 |
| User 6 Gaze Puzzle 2A | 0.496 | 0.298 | 1.313 | 11 | 25.673 | <0.01 |
| User 6 Gaze Puzzle 2B | 0.419 | -0.012 | 1.235 | 22 | 75.3 | <0.01 |
| User 6 Gaze Puzzle 3 | 0.325 | 0.647 | 1.039 | 10 | 22.571 | 0.012 |
| User 6 Gaze Puzzle 4A | 0.493 | 0.363 | 1.049 | 11 | 7.549 | 0.753 |
| User 6 Gaze Puzzle 4B | 0.433 | 0.442 | 1.137 | 12 | 27.678 | <0.01 |
| User 1 Speech Puzzle 1 | 0.039 | 0.472 | 0.877 | 6 | 0.842 | 0.991 |
| User 1 Speech Puzzle 2A | 0.106 | 0.33 | 1.311 | 7 | 7.2 | 0.408 |
| User 1 Speech Puzzle 2B | -0.133 | 0.879 | 1.042 | 7 | 13 | 0.072 |
| User 1 Speech Puzzle 3 | 0.133 | 0.518 | 1.298 | 8 | 12.281 | 0.139 |
| User 1 Speech Puzzle 4A | 0.286 | 0.231 | 1.255 | 5 | 15.194 | <0.01 |
| User 1 Speech Puzzle 4B | 0.377 | -0.243 | 1.381 | 6 | 7.091 | 0.313 |
| User 2 Speech Puzzle 1 | 0.603 | -0.114 | 1.176 | 7 | 5.259 | 0.628 |
| User 2 Speech Puzzle 2A | 0.307 | 0.075 | 0.837 | 11 | 16.375 | 0.128 |
| User 2 Speech Puzzle 2B | 0.407 | -0.132 | 1.052 | 7 | 11.556 | 0.116 |
| User 2 Speech Puzzle 3 | 0.546 | -0.353 | 1.093 | 7 | 9.075 | 0.247 |
| User 2 Speech Puzzle 4A | 0.308 | 0.272 | 0.81 | 10 | 20.058 | 0.029 |
| User 2 Speech Puzzle 4B | 0.547 | 0.174 | 1.027 | 13 | 24.779 | 0.025 |
| User 3 Speech Puzzle 1 | 0.221 | 0.369 | 1.004 | 6 | 6.054 | 0.417 |
| User 3 Speech Puzzle 2A | -0.27 | 0.782 | 0.72 | 9 | 12.154 | 0.205 |

| Participants | $A_{Ln}$ | m | s | df | $X^2$ | P |
|---|---|---|---|---|---|---|
| User 3 Speech Puzzle 2B | -0.065 | 0.601 | 0.954 | 6 | 10.714 | 0.098 |
| User 3 Speech Puzzle 3 | 0.27 | 0.423 | 1.182 | 8 | 10.915 | 0.207 |
| User 3 Speech Puzzle 4A | 0.379 | 0.434 | 0.967 | 13 | 33.365 | <0.01 |
| User 3 Speech Puzzle 4B | 0.147 | 0.280 | 1.248 | 4 | 5.647 | 0.227 |
| User 4 Speech Puzzle 1 | 0.169 | 0.702 | 1.182 | 15 | 27.167 | 0.027 |
| User 4 Speech Puzzle 2A | 0.337 | 0.309 | 1.043 | 12 | 22.652 | 0.031 |
| User 4 Speech Puzzle 2B | 0.169 | 0.323 | 1.161 | 10 | 22.956 | 0.011 |
| User 4 Speech Puzzle 3 | 0.340 | 0.573 | 1.359 | 10 | 15.809 | 0.105 |
| User 4 Speech Puzzle 4A | 0.354 | 0.663 | 1.256 | 15 | 38.412 | <0.01 |
| User 4 Speech Puzzle 4B | 0.310 | 0.403 | 1.164 | 9 | 8.646 | 0.471 |
| User 5 Speech Puzzle 1 | 0.468 | 0.176 | 1.189 | 5 | 2 | 0.849 |
| User 5 Speech Puzzle 2A | 0.653 | 0.198 | 1.203 | 8 | 8 | 0.434 |
| User 5 Speech Puzzle 2B | 0.675 | 0.027 | 1.211 | 8 | 12.657 | 0.124 |
| User 5 Speech Puzzle 3 | 0.809 | 0.041 | 1.379 | 5 | 5.903 | 0.316 |
| User 5 Speech Puzzle 4A | 0.450 | 0.550 | 1.028 | 20 | 45.554 | <0.01 |
| User 5 Speech Puzzle 4B | 0.708 | 0.444 | 1.197 | 4 | 2 | 0.736 |
| User 6 Speech Puzzle 1 | 0.026 | 0.343 | 1.153 | 5 | 11 | 0.051 |
| User 6 Speech Puzzle 2A | 0.304 | -0.005 | 0.958 | 6 | 3.059 | 0.801 |
| User 6 Speech Puzzle 2B | 0.127 | -0.049 | 1.128 | 7 | 10.522 | 0.161 |
| User 6 Speech Puzzle 3 | 0.349 | -0.319 | 1.18 | 4 | 3.176 | 0.529 |
| User 6 Speech Puzzle 4A | 0.358 | -0.242 | 1.311 | 5 | 2.2 | 0.821 |
| User 6 Speech Puzzle 4B | 0.710 | -0.718 | 1.444 | 4 | 3.455 | 0.485 |

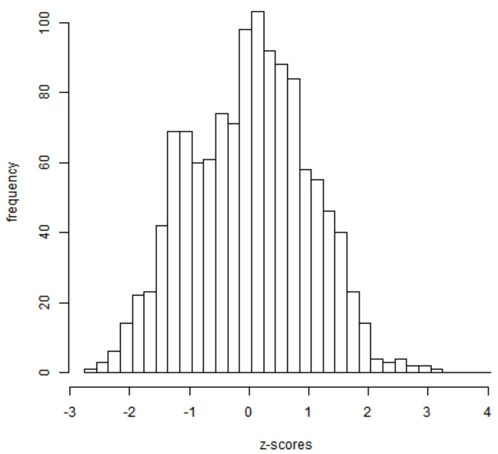
**4 Histograms of normalized log-transformed latencies; each diagram contains latencies of each puzzle per interaction**



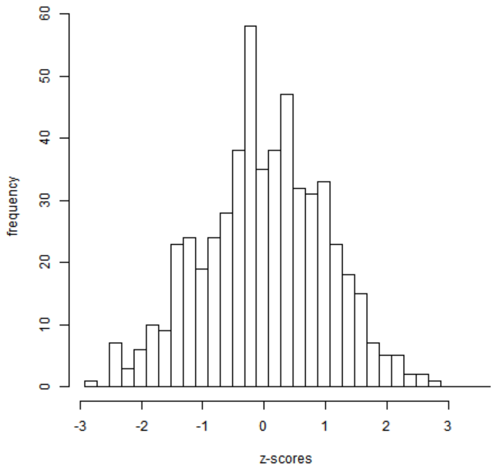

**Touch Puzzle 1**



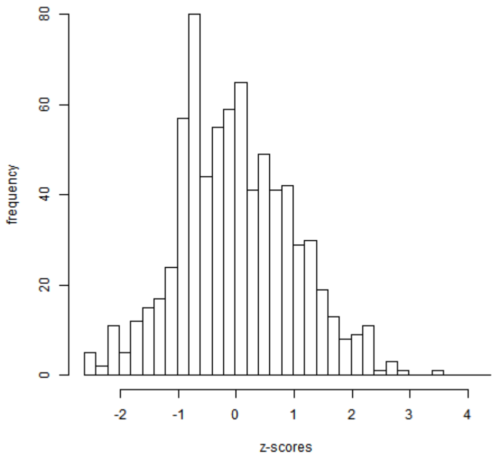

**Touch Puzzle 2A**



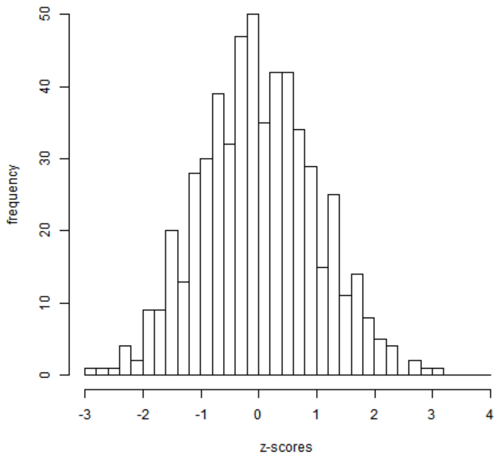

**Touch Puzzle 2B**



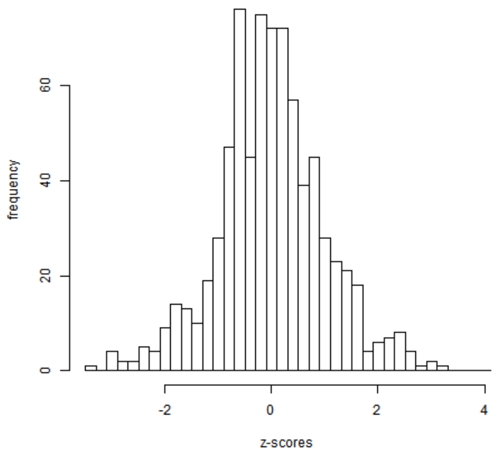

**Touch Puzzle 3**



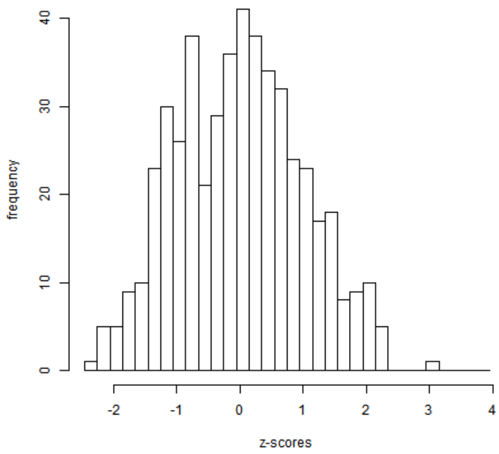

**Touch Puzzle 4A**



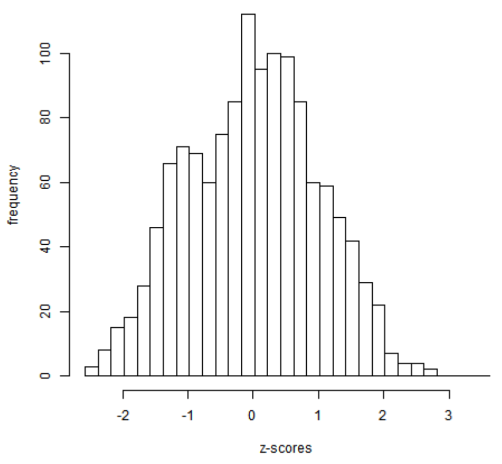

**Touch Puzzle 4B**



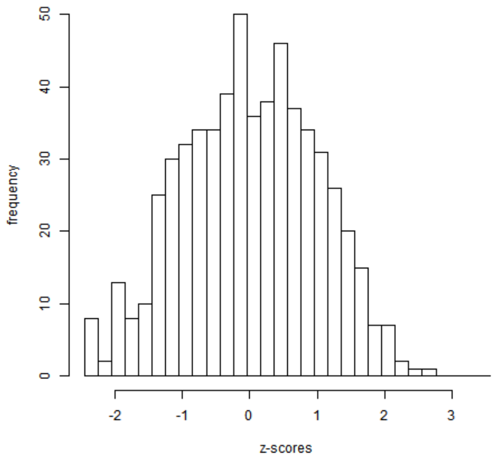

**Gaze Puzzle 1**



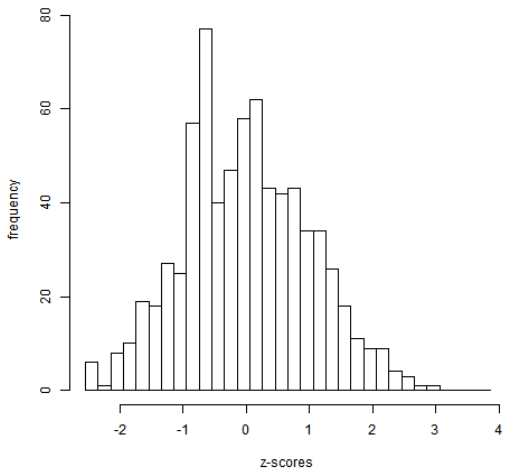

**Gaze Puzzle 2A**



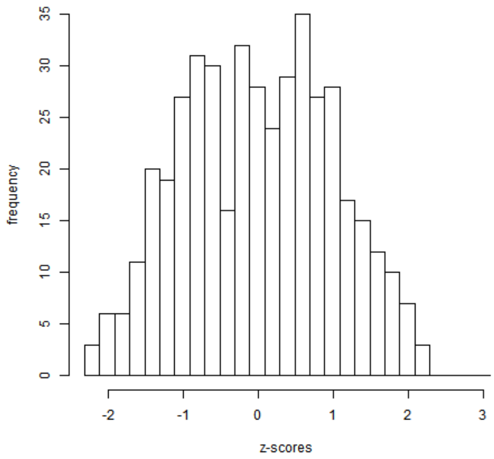

**Gaze Puzzle 2B**



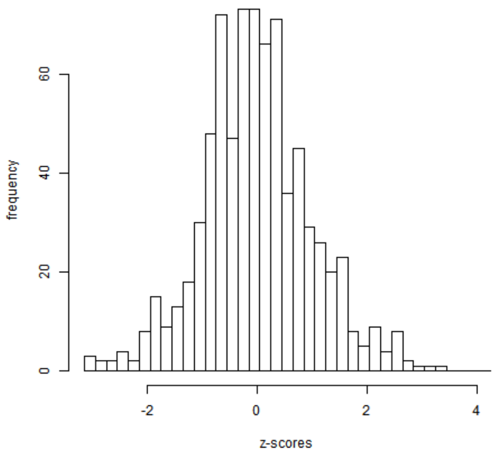

**Gaze Puzzle 3**



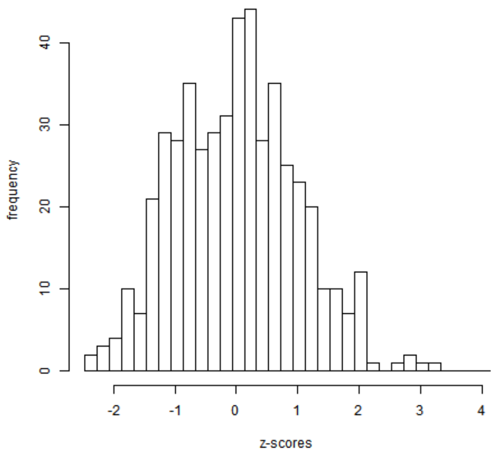

**Gaze Puzzle 4A**



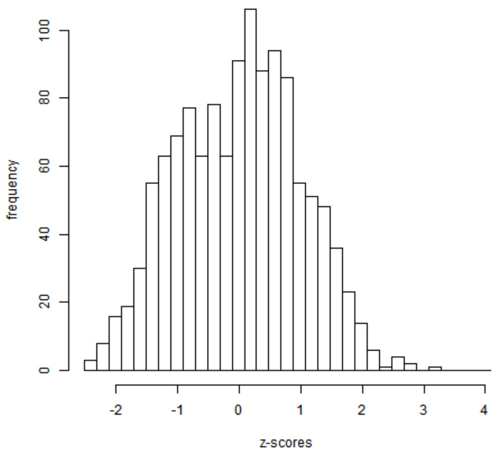

**Gaze Puzzle 4B**



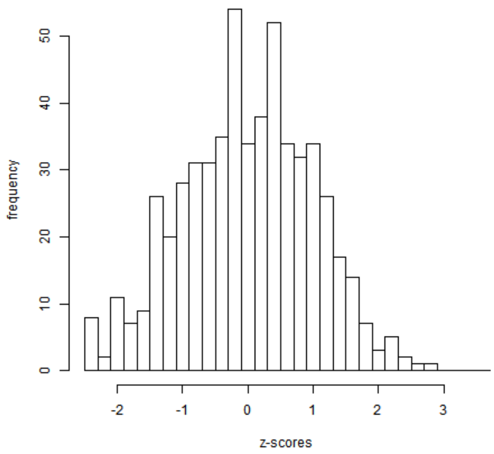

**Speech Puzzle 1**



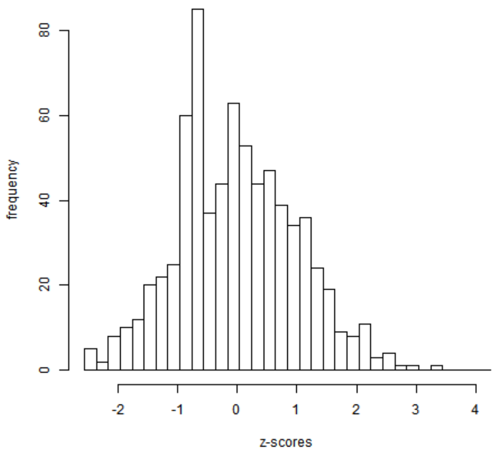

**Speech Puzzle 2A**



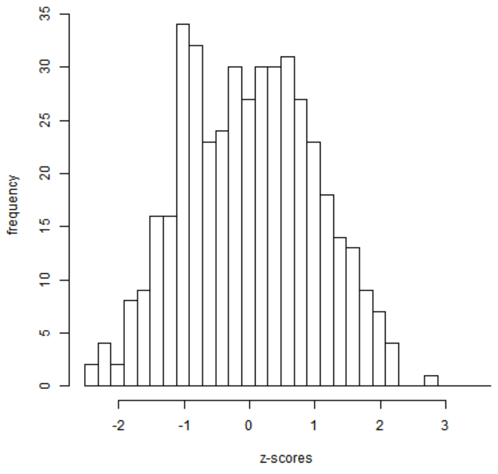

**Speech Puzzle 2B**




**Speech Puzzle 3**

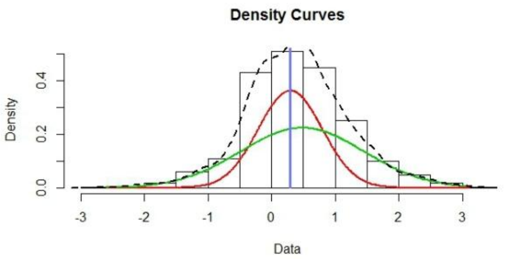



**Speech Puzzle 4A**

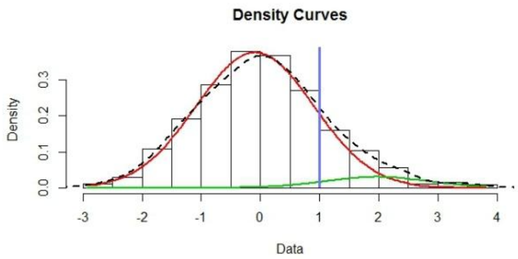



**Speech Puzzle 4B**

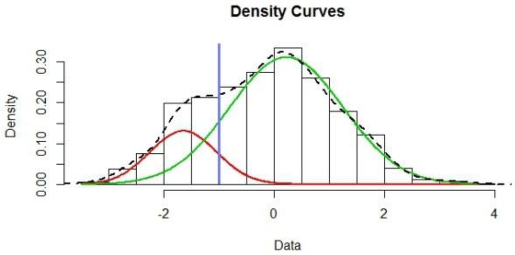
**5 Manual Thresholds; including Threshold (bold vertical line), Gaussian density components in the mixture distribution (solid lines), and estimated density function (dashed lines)**



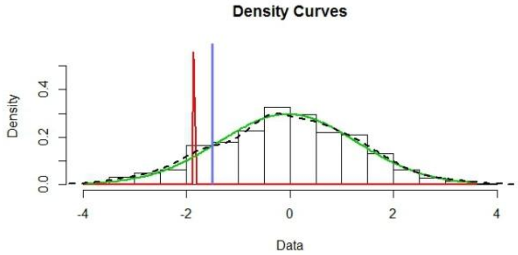

**Touch Puzzle 1**



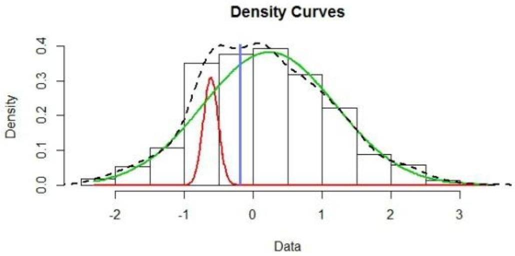

**Touch Puzzle 2A**



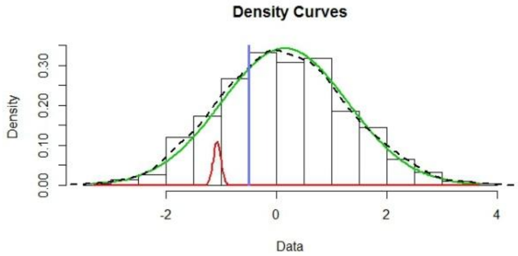

**Touch Puzzle 2B**



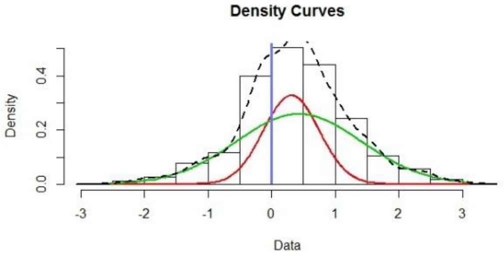

**Touch Puzzle 3**



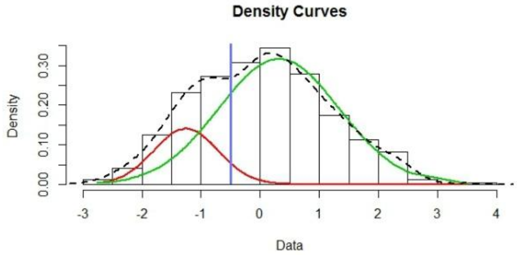

**Touch Puzzle 4A**



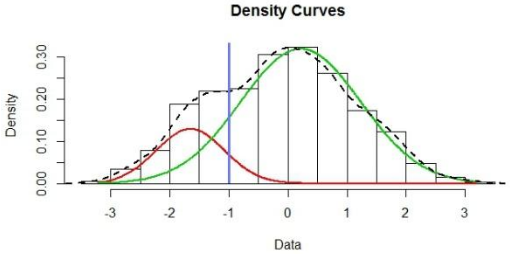

**Touch Puzzle 4B**



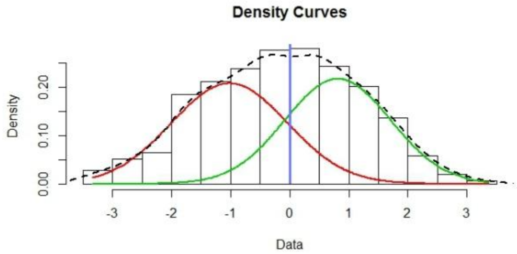

**Gaze Puzzle 1**



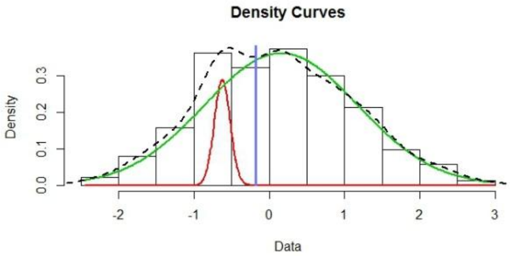

**Gaze Puzzle 2A**



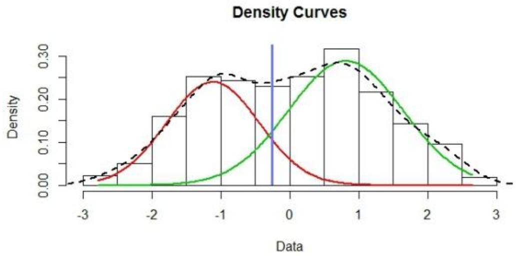

**Gaze Puzzle 2B**



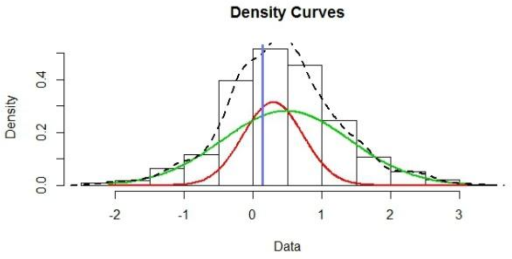

**Gaze Puzzle 3**



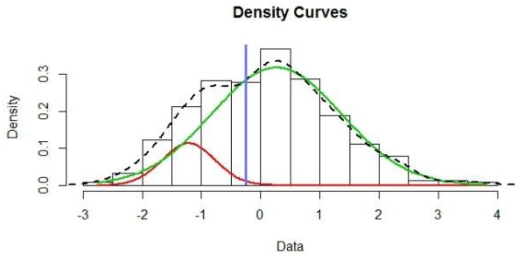

**Gaze Puzzle 4A**



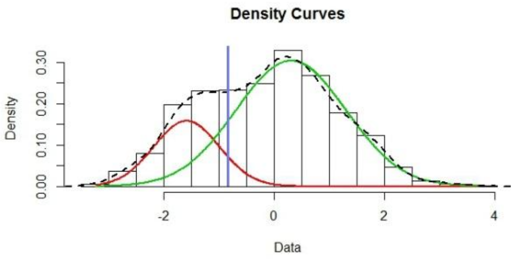

**Gaze Puzzle 4B**



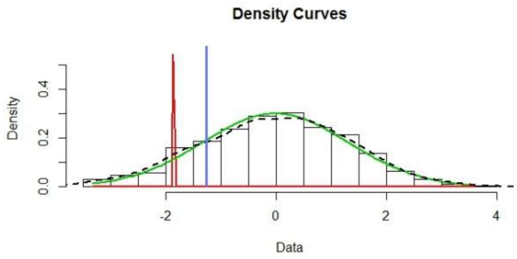

**Speech Puzzle 1**



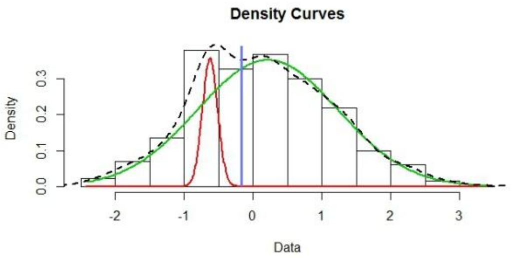

**Speech Puzzle 2A**



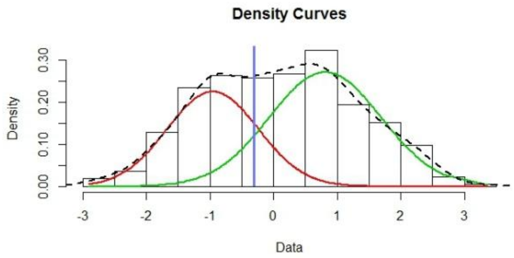

**Speech Puzzle 2B**




**Speech Puzzle 3**




**Speech Puzzle 4A**




**Speech Puzzle 4B**